

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79234 A2

(51) International Patent Classification⁷: **C07H**

(74) Agent: **FIELD, Gisela, M.**; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(21) International Application Number: **PCT/US01/12303**

(22) International Filing Date: **16 April 2001 (16.04.2001)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:
60/197,188 **14 April 2000 (14.04.2000)** **US**

(71) Applicant (for all designated States except US): **GENAIS-
SANCE PHARMACEUTICALS, INC.** [US/US]; Five
Science Park, New Haven, CT 06511 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KLIEM, Stephanie, E.** [DE/US]; 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **LEE, Helen, H.** [US/US]; 26 Rivendell Drive, Shelton, CT 06484 (US).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **HAPLOTYPES OF THE APOE GENE**

(57) Abstract: Novel single nucleotide polymorphisms in the human apolipoprotein E (APOE) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the APOE gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the APOE gene in an individual are also disclosed. Polynucleotides containing one or more of the APOE polymorphisms disclosed herein are also described.

WO 01/79234 A2

HAPLOTYPES OF THE APOE GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/197,188 filed April 14, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human apolipoprotein E (APOE) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical trials or their early withdrawal from the

market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* 15:1249-52; Kleyn PW et al. 1998 *Science* 281: 1820-21; Kola I 1999 *Curr Opin Biotech* 10:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* 66:445-7; Marshall, E 1999 *Science* 284:406-7; Judson R et al. 2000 *Pharmacogenomics* 1:1-12; Roses AD 2000 *Nature* 405:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* 19:216-7; Wang DG et al 1998 *Science* 280:1077-82; Chakravarti A 1999 *Nat Genet* 21:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* 4:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* 5:538-43; Davidson S 2000 *Nature Biotech* 18:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* 320:987-90; Dahl BS 1997 *Acta Psychiatr Scand* 96 (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* 97:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease is the apolipoprotein E (APOE) gene or its encoded product. APOE is a plasma protein that mediates the transport and uptake of cholesterol and lipid by binding to various cellular receptors, including low-density lipoprotein (LDL) receptor (Wilson et al., *Science* 1991; 252:1817-1822). APOE is one of the major

constituents of chylomicrons and very-low-density lipoprotein (VLDL) remnants, and functions as a ligand for the uptake of these particles by the liver (de Knijff et al., *Hum. Mutat.* 1994; 4:178-194). APOE is present in high concentrations in intestinal fluid where it plays a role in cholesterol redistribution from cells with excess cholesterol to those requiring cholesterol. APOE is also involved in the repair-response to tissue injury, immunoregulation and modulation of cell growth, and differentiation (Mahley, *Science* 1988; 240:622-630).

The apolipoprotein E gene is located on chromosome 19q13.2 and contains 4 exons that encode a 317 amino acid protein. Reference sequences for the APOE gene (Genaissance Reference No. 1515984; SEQ ID NO:1), coding sequence (GenBank Accession No:NM_000041.1), and protein are shown in Figures 1, 2 and 3, respectively.

Defective APOE binding to low density lipoprotein receptors is associated with familial type III hyperlipoproteinemia, a genetic disorder characterized by elevated plasma cholesterol levels and accelerated coronary artery disease (Mahley et al., *supra*). Polymorphic variants of APOE are also associated with familial dysbetalipoproteinemia characterized by elevated plasma cholesterol and triglyceride levels and an increased risk for atherosclerosis (de Knijff et al., *supra*).

APOE exists in three common polymorphic variants (APOE2,-E3,-E4) that differ in their amino acid sequences at two positions, residue 112 (called A site) and residue 158 (called B site). APOE4 is associated with the late-onset familial and sporadic forms of Alzheimer's disease (AD). It has been found that an increasing amount of this allele correlates with a 20-90% increased risk for AD and a decreased mean age of onset (Corder et al., *Science* 1993; 261:921-923). It has also been found that this variant binds to the amyloid beta peptide and microtubule-associated protein (tau), both of which are major components of the neuritic plaque in AD, and possibly affect microtubule function in the Alzheimer brain (Strittmatter and Roses, *Proc. Natl. Acad. Sci. U.S.A* 1995; 92:4725-4727).

Two polymorphisms in the APOE gene have been reported in the Human Gene Mutation Database (HGMD) that are associated with APOE deficiency. These correspond to a polymorphism of a thymine or cytosine at a position corresponding to nucleotide position 21250 and a polymorphism of cytosine or thymine at position corresponding to nucleotide position 21388 respectively, in Figure 1. These two polymorphisms result in amino acid variations of cysteine or arginine at amino acid position 130 and an arginine or cysteine at amino acid position 176, respectively in Figure 3. Because of the potential for variation in the APOE gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the APOE gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of APOE as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 14 novel polymorphic sites in the APOE

gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 17936 (PS1), 18393 (PS2), 18445 (PS3), 18476 (PS4), 18489 (PS5), 18544 (PS6), 18592 (PS7), 19203 (PS8), 19221 (PS9), 19311 (PS10), 21226 (PS11), 21360 (PS13), 21484 (PS15) and 21818 (PS16). The polymorphisms at these sites are cytosine or thymine at PS1, cytosine or thymine at PS2, guanine or adenine at PS3, cytosine or guanine at PS4, guanine or adenine at PS5, guanine or adenine at PS6, cytosine or adenine at PS7, thymine or cytosine at PS8, adenine or guanine at PS9, guanine or adenine at PS10, cytosine or adenine at PS11, cytosine or guanine at PS13, guanine or adenine at PS15 and cytosine or thymine at PS16. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified sites at nucleotide positions 21250 (PS12) and 21388 (PS14), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS16 in the APOE gene, which are shown below in Tables 5 and 4, respectively. Each of these APOE haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the APOE gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the APOE gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16 in both copies of the APOE gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel APOE polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel APOE polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-PS16. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the APOE gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the APOE gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's APOE gene is defined by one of the APOE haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's APOE gene are defined by one of the APOE haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the APOE haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in

the discovery and development of drugs for treating diseases associated with APOE activity, e.g., Familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate APOE as a candidate target for treating a specific condition or disease predicted to be associated with APOE activity. Determining for a particular population the frequency of one or more of the individual APOE haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue APOE as a target for treating the specific disease of interest. In particular, if variable APOE activity is associated with the disease, then one or more APOE haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed APOE haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable APOE activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any APOE haplotype or haplotype pair, apply the information derived from detecting APOE haplotypes in an individual to decide whether modulating APOE activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting APOE to treat a specific condition or disease predicted to be associated with APOE activity. For example, detecting which of the APOE haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent APOE isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular APOE haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the APOE gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with APOE activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the APOE haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute APOE haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of an APOE haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any APOE haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association

between a trait and an APOE genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the APOE genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the APOE genotype or haplotype in a reference population. A higher frequency of the APOE genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the APOE genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the APOE haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the APOE gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, adenine at PS5, adenine at PS6, adenine at PS7, cytosine at PS8, guanine at PS9, adenine at PS10, adenine at PS11, guanine at PS13, adenine at PS15 and thymine at PS16. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of cytosine at PS12 and thymine at PS14.

A particularly preferred polymorphic variant is an isogene of the APOE gene. An APOE isogene of the invention comprises cytosine or thymine at PS1, cytosine or thymine at PS2, guanine or adenine at PS3, cytosine or guanine at PS4, guanine or adenine at PS5, guanine or adenine at PS6, cytosine or adenine at PS7, thymine or cytosine at PS8, adenine or guanine at PS9, guanine or adenine at PS10, cytosine or adenine at PS11, thymine or cytosine at PS12, cytosine or guanine at PS13, cytosine or thymine at PS14, guanine or adenine at PS15 and cytosine or thymine at PS16. The invention also provides a collection of APOE isogenes, referred to herein as an APOE genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for an APOE cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of cytosine at a position corresponding to nucleotide 13, guanine at a position corresponding to nucleotide 31, adenine at a position corresponding to nucleotide 364, guanine at a position corresponding to nucleotide 498, and adenine at a position corresponding to nucleotide 622. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of cytosine at a position corresponding to nucleotide 388 and thymine at a position corresponding to nucleotide 526. A particularly preferred

polymorphic cDNA variant comprises the coding sequence of an APOE isogene defined by haplotypes 1-22.

Polynucleotides complementary to these APOE genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the APOE-gene will be useful in studying the expression and function of APOE, and in expressing APOE protein for use in screening for candidate drugs to treat diseases related to APOE activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express APOE for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the APOE protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of arginine at a position corresponding to amino acid position 5, alanine at a position corresponding to amino acid position 11, methionine at a position corresponding to amino acid position 122, and methionine at a position corresponding to amino acid position 208. In some embodiments, the polymorphic variant also comprises at least one variant amino acid selected from the group consisting of arginine at a position corresponding to amino acid position 130 and cysteine at a position corresponding to amino acid position 176. A polymorphic variant of APOE is useful in studying the effect of the variation on the biological activity of APOE as well as on the binding affinity of candidate drugs targeting APOE for the treatment of familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease.

The present invention also provides antibodies that recognize and bind to the above polymorphic APOE protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the APOE polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the APOE isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against APOE protein, and for testing the efficacy of therapeutic agents and compounds for familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the APOE gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the APOE gene in a reference population. In a preferred embodiment, the computer system is capable of producing a

display showing APOE haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the APOE gene (Genaissance Reference No. 1515984; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:76 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the APOE coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the APOE protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the APOE gene. As described in more detail below, the inventors herein discovered 22 isogenes of the APOE gene by characterizing the APOE gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The APOE isogenes present in the human reference population are defined by haplotypes for 16 polymorphic sites in the APOE gene, 14 of which are believed to be novel. The APOE polymorphic sites identified by the inventors are referred to as PS1-PS16 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16. Using the genotypes identified in the Index Repository for PS1-PS16 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the APOE gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the APOE gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether APOE is a suitable target for drugs to treat Familial dysbetalipoproteinemia, type III hyperlipoproteinemia, atherosclerosis, and Alzheimer's disease, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene - A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype - An unphased 5' to 3' sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype - The unphased 5' to 3' sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype - The unphased 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping - A process for determining a genotype of an individual.

Haplotype - A 5' to 3' sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype - The 5' to 3' sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype - The 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair - The two haplotypes found for a locus in a single individual.

Haplotyping - A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform - A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene - One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated - As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to

absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring - A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair - The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased - As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) - A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant - A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism - The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data - Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database - A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide - A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group - A group of individuals sharing a common ethnogeographic origin.

Reference Population - A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) - Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject - A human individual whose genotypes or haplotypes or response to treatment or

disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased - As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the APOE gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel APOE polymorphisms and haplotypes identified herein.

The compositions comprise at least one APOE genotyping oligonucleotide. In one embodiment, an APOE genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of an APOE polynucleotide, i.e., an APOE isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-APOE polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the APOE gene using the polymorphism information provided herein in conjunction with the known sequence information for

the APOE gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5' end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in *PCR Protocols, A Guide to Methods and Applications*, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y=

T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting APOE gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

GGCGTGAYTACCGCC (SEQ ID NO:4) and its complement,
 GGACGTCYTTCCCA (SEQ ID NO:5) and its complement,
 AGCTCAGRGGCCTCT (SEQ ID NO:6) and its complement,
 CTGGGAASCCCTGGC (SEQ ID NO:7) and its complement,
 GCCTCCARGTAGTCT (SEQ ID NO:8) and its complement,
 CGGGGGTRAGGCAAG (SEQ ID NO:9) and its complement,
 GCAGAGAMGACCCGA (SEQ ID NO:10) and its complement,
 GGTTCYGGGGCTGC (SEQ ID NO:11) and its complement,
 GCTGGTCRCATTCCT (SEQ ID NO:12) and its complement,
 CATTCAGRCAGACCC (SEQ ID NO:13) and its complement,
 GGCCCGMTGGGCGC (SEQ ID NO:14) and its complement,
 AGCGGCTSCTCCGCG (SEQ ID NO:15) and its complement,
 GGGCCGCGTGGGGC (SEQ ID NO:16) and its complement, and
 CACTGAAYGCCGAAG (SEQ ID NO:17) and its complement.

A preferred ASO primer for detecting APOE gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

ATTACAGGCGTGAYT (SEQ ID NO:18); GCTGGGGGCGGTART (SEQ ID NO:19);
 GGTGAAGGACGTCYT (SEQ ID NO:20); GGCTCCTGGGGAARG (SEQ ID NO:21);
 GGGATGAGCTCAGRG (SEQ ID NO:22); CTTTCTAGAGGCCYC (SEQ ID NO:23);
 GGGACCCTGGGAASC (SEQ ID NO:24); CTGGAGGCCAGGGST (SEQ ID NO:25);
 CCCCTGGCCTCCARG (SEQ ID NO:26); CTCCTGAGACTACYT (SEQ ID NO:27);
 GAGGAGCGGGGGTRA (SEQ ID NO:28); CTGCTGCTTGCCTYA (SEQ ID NO:29);
 TGGGCAGCAGAGAMG (SEQ ID NO:30); AGCGGGTCGGGTCKT (SEQ ID NO:31);
 GATGAAGGTTCTGYG (SEQ ID NO:32); AGCAACGCAGCCRC (SEQ ID NO:33);
 TGCCTTGCTGGTCRC (SEQ ID NO:34); CCTGCCAGGAATGYG (SEQ ID NO:35);
 TGGCCCCATTCAGRC (SEQ ID NO:36); GGCCCAGGGTCTGYC (SEQ ID NO:37);
 GGCGCAGGCCCGMT (SEQ ID NO:38); ATGTCCGCGCCCAKC (SEQ ID NO:39);
 TGCCTAAGCGGCTSC (SEQ ID NO:40); CGGCATCGCGGAGSA (SEQ ID NO:41);
 GGAACAGGGCCGCGRT (SEQ ID NO:42); GTGGCGGCCCGCAYG (SEQ ID NO:43);
 GACAATCACTGAAYG (SEQ ID NO:44); and TGCAGGCTTCGGCRT (SEQ ID NO:45).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting APOE gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

ACAGGCGTGA (SEQ ID NO: 46); GGGGGCGGTA (SEQ ID NO: 47);
 GAAGGACGTC (SEQ ID NO: 48); TCCTGGGGAA (SEQ ID NO: 49);
 ATGAGCTCAG (SEQ ID NO: 50); TCTAGAGGCC (SEQ ID NO: 51);
 ACCCTGGGAA (SEQ ID NO: 52); GAGGCCAGGG (SEQ ID NO: 53);
 CTGGCCTCCA (SEQ ID NO: 54); CTGAGACTAC (SEQ ID NO: 55);
 GAGCGGGGGT (SEQ ID NO: 56); CTGCTTGCCT (SEQ ID NO: 57);
 GCAGCAGAGA (SEQ ID NO: 58); GGGTCGGGTC (SEQ ID NO: 59);
 GAAGGTTCTG (SEQ ID NO: 60); AACGCAGCCC (SEQ ID NO: 61);
 GTTGCTGGTC (SEQ ID NO: 62); GCCAGGAATG (SEQ ID NO: 63);
 CCCCATTCTG (SEQ ID NO: 64); CCAGGGTCTG (SEQ ID NO: 65);
 GCAGGCCCGG (SEQ ID NO: 66); TCCGCGCCCA (SEQ ID NO: 67);
 GTAAGCGGCT (SEQ ID NO: 68); CATCGCGGAG (SEQ ID NO: 69);
 ACAGGGCCGC (SEQ ID NO: 70); GCGGCCCGCA (SEQ ID NO: 71);
 AATCACTGAA (SEQ ID NO: 72); and AGGCTTCGGC (SEQ ID NO: 73).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

APOE genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized APOE genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the APOE gene in an individual. As used herein, the terms "APOE genotype" and "APOE haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the APOE gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic

acid sample comprising the two copies of the APOE gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16 in the two copies to assign an APOE genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair at one or more of the polymorphic sites selected from the group consisting of PS12 and PS14 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS16.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the APOE gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5' and 3' untranslated regions. If an APOE gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the APOE gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16 in that copy to assign an APOE haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the APOE gene or fragment such as one of the methods described above for preparing APOE isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two APOE gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional APOE clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the APOE gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at one or more polymorphic sites selected from the group consisting of PS12 and PS14. In a particularly preferred embodiment, the nucleotide at each of PS1-PS16 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the APOE haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's APOE gene, the phased sequence of nucleotides present at each of PS1-PS16. The present invention also contemplates that typically only a subset of PS1-PS16

will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysedale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, an APOE haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16 in each copy of the APOE gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS16 in each copy of the APOE gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the APOE gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188),

ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the APOE gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, *P. Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism

(SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in *Molecular Diagnosis of Genetic Diseases*, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruano et al., *Nucl. Acids Res.* 17:8392, 1989; Ruano et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's APOE haplotype pair is predicted from its APOE genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying an APOE genotype for the individual at two or more APOE polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing APOE haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the APOE haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the

population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n = \log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., *Principles of Population Genomics*, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting an APOE haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example,

CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of an APOE genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel APOE polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for APOE genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and an APOE genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular APOE genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that APOE genotype, haplotype or haplotype pair. Preferably, the APOE genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting APOE or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all

of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and an APOE genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the APOE gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and APOE genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their APOE genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health

Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the APOE gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between APOE haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the APOE gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of APOE genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the APOE gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the APOE gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying APOE genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the APOE gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant APOE gene is

identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16, and may also comprise one or more additional polymorphisms selected from the group consisting of cytosine at PS12 and thymine at PS14. Similarly, the nucleotide sequence of a variant fragment of the APOE gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the APOE gene (or other reported APOE sequences) or to portions of the reference sequence (or other reported APOE sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, adenine at PS5, adenine at PS6, adenine at PS7, cytosine at PS8, guanine at PS9, adenine at PS10, adenine at PS11, guanine at PS13, adenine at PS15 and thymine at PS16. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the APOE gene which is defined by any one of haplotypes 1-22 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the APOE gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

APOE isogenes may be isolated using any method that allows separation of the two "copies" of the APOE gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides APOE genome anthologies, which are collections of APOE isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. An APOE genome anthology may comprise individual APOE isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and

the like. Alternatively, two or more groups of the APOE isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred APOE genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded APOE protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant APOE sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as *E. coli*, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the APOE gene will produce APOE mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of an APOE cDNA comprising a nucleotide sequence which is a polymorphic variant of the APOE reference coding sequence shown in Figure 2. Thus, the invention also provides APOE mRNAs and

corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of cytosine at a position corresponding to nucleotide 13, guanine at a position corresponding to nucleotide 31, adenine at a position corresponding to nucleotide 364, guanine at a position corresponding to nucleotide 498, and adenine at a position corresponding to nucleotide 622, and may also comprise one or more additional polymorphisms selected from the group consisting of cytosine at a position corresponding to nucleotide 388, thymine at a position corresponding to nucleotide 526. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an APOE isogene defined by haplotypes 1-22. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized APOE cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of an APOE gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the APOE polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the APOE gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the APOE genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular APOE protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the APOE isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular APOE isogene. Expression of an APOE isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter,

introns, enhancers, 3' untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions -10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of APOE mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of APOE mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2' O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference APOE amino acid sequence shown in Figure 3. The location of a variant amino acid in an APOE polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). An APOE protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 5, alanine at a position corresponding to amino acid position 11, methionine at a position corresponding to amino acid position 122, and methionine at a position corresponding to amino acid position 208, and may also comprise one or more additional variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 130 and cysteine at a position corresponding to amino acid position 176. The invention specifically excludes amino acid sequences identical to those previously identified for APOE, including SEQ ID NO:3, and previously described fragments thereof. APOE protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, an APOE protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2.
Polymorphic
Variant

Novel Polymorphic Variants of APOE
Amino Acid Position and Identities

Number	5	11	122	130	176	208
1	W	T	L	C	R	M
2	W	T	L	C	C	M
3	W	T	L	R	R	M
4	W	T	L	R	C	M
5	W	T	M	C	R	V
6	W	T	M	C	R	M
7	W	T	M	C	C	V
8	W	T	M	C	C	M
9	W	T	M	R	R	V
10	W	T	M	R	R	M
11	W	T	M	R	C	V
12	W	T	M	R	C	M
13	W	A	L	C	R	V
14	W	A	L	C	R	M
15	W	A	L	C	C	V
16	W	A	L	C	C	M
17	W	A	L	R	R	V
18	W	A	L	R	R	M
19	W	A	L	R	C	V
20	W	A	L	R	C	M
21	W	A	M	C	R	V
22	W	A	M	C	R	M
23	W	A	M	C	C	V
24	W	A	M	C	C	M
25	W	A	M	R	R	V
26	W	A	M	R	R	M
27	W	A	M	R	C	V
28	W	A	M	R	C	M
29	R	T	L	C	R	V

Table 2.
Polymorphic
Variant

Novel Polymorphic Variants of APOE(Cont)
Amino Acid Position and Identities

Number	5	11	122	130	176	208
30	R	T	L	C	R	M
31	R	T	L	C	C	V
32	R	T	L	C	C	M
33	R	T	L	R	R	V
34	R	T	L	R	R	M
35	R	T	L	R	C	V
36	R	T	L	R	C	M
37	R	T	M	C	R	V
38	R	T	M	C	R	M
39	R	T	M	C	C	V
40	R	T	M	C	C	M
41	R	T	M	R	R	V
42	R	T	M	R	R	M
43	R	T	M	R	C	V
44	R	T	M	R	C	M
45	R	A	L	C	R	V
46	R	A	L	C	R	M
47	R	A	L	C	C	V
48	R	A	L	C	C	M
49	R	A	L	R	R	V
50	R	A	L	R	R	M
51	R	A	L	R	C	V
52	R	A	L	R	C	M
53	R	A	M	C	R	V
54	R	A	M	C	R	M
55	R	A	M	C	C	V
56	R	A	M	C	C	M
57	R	A	M	R	R	V
58	R	A	M	R	R	M
59	R	A	M	R	C	V
60	R	A	M	R	C	M

The invention also includes APOE peptide variants, which are any fragments of an APOE protein variant that contain one or more of the amino acid variations shown in Table 2. An APOE peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such APOE peptide variants may be useful as antigens to generate antibodies specific for one of the above APOE isoforms. In addition, the APOE peptide variants may be useful in drug screening assays.

An APOE variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant APOE genomic and cDNA sequences as described above. Alternatively, the APOE protein variant may be isolated from a biological sample of an individual having an APOE isogene which encodes the variant protein. Where the sample contains two different APOE isoforms (i.e., the individual has different APOE isogenes), a particular APOE isoform of the

invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular APOE isoform but does not bind to the other APOE isoform.

The expressed or isolated APOE protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the APOE protein as discussed further below. APOE variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant APOE gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric APOE protein. The non-APOE portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the APOE and non-APOE portions so that the APOE protein may be cleaved and purified away from the non-APOE portion.

An additional embodiment of the invention relates to using a novel APOE protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known APOE protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The APOE protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to an APOE variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the APOE protein(s) of interest and then washed. Bound APOE protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel APOE protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the APOE protein.

In yet another embodiment, when a particular APOE haplotype or group of APOE haplotypes encodes an APOE protein variant with an amino acid sequence distinct from that of APOE protein isoforms encoded by other APOE haplotypes, then detection of that particular APOE haplotype or group of APOE haplotypes may be accomplished by detecting expression of the encoded APOE protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel APOE variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The APOE protein or peptide variant used to generate the

antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the APOE protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the APOE isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the APOE protein variant from solution as well as react with APOE protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect APOE protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel APOE protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the APOE protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or

those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, *Nature*, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: *Laboratory Techniques in Biochemistry and Molecular Biology*, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, *Science*, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 *Proc. Natl. Acad. Sci. USA* 86:10029).

Effect(s) of the polymorphisms identified herein on expression of APOE may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the APOE gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into APOE protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired APOE isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the APOE isogene is introduced into a cell in such a way that it recombines with the endogenous APOE gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired APOE gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the APOE isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the APOE isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant APOE gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill

in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the APOE isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human APOE isogene and producing human APOE protein can be used as biological models for studying diseases related to abnormal APOE expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel APOE isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel APOE isogenes; an antisense oligonucleotide directed against one of the novel APOE isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel APOE isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel APOE isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors

relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the APOE gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The APOE polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the APOE gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the APOE gene were amplified using either PCR primer pairs represented below by providing the nucleotide positions of their initial and final nucleotides, which

correspond to positions in Figure 1, or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5' end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5'-TGTAACGACGGCCAGT-3' (SEQ ID NO:74) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5'-AGGAAACAGCTATGACCAT-3' (SEQ ID NO:75). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	17862-17883	complement of 18454-18433	593 nt
Fragment 2	18186-18208	complement of 18660-18637	475 nt
Fragment 3	19027-19049	complement of 19411-19390	385 nt
Fragment 4	20083-20103	complement of 20694-20670	612 nt
Fragment 5	20910-20931	complement of 21585-21566	676 nt
Fragment 6	21103-21124	complement of 21715-21696	613 nt
Fragment 7	21383-21403	complement of 22149-22128	767 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 μ l
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 μ l
100 ng of human genomic DNA	= 1 μ l
10 mM dNTP	= 0.4 μ l
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 μ l
Forward Primer (10 μ M)	= 0.4 μ l
Reverse Primer (10 μ M)	= 0.4 μ l
Water	= 6.6 μ l

Amplification profile:

97°C - 2 min.	1 cycle
97°C - 15 sec.	} 10 cycles
70°C - 45 sec.	
72°C - 45 sec.	
97°C - 15 sec.	} 35 cycles
64°C - 45 sec.	
72°C - 45 sec.	

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 μ l 384 well unfilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 μ l of

distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1, or the appropriate universal 'tail' sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	Tailed sequence	
Fragment 2	18209-18228	complement of 18579-18560
Fragment 3	19061-19080	complement of 19359-19340
Fragment 4	20134-20153	complement of 20650-20629
Fragment 5	20979-20998	complement of 21403-21384
Fragment 6	21132-21152	complement of 21626-21607
Fragment 7	21647-21666	complement of 22128-22109

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the APOE gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the APOE Gene						
Polymorphic Site Number	PolyId ^a	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	7781951	17936	C	T		
PS2	1516004	18393	C	T		
PS3	1516006	18445	G	A		
PS4	1516008	18476	C	G		
PS5	1516010	18489	G	A		
PS6	1516012	18544	G	A		
PS7	1516014	18592	C	A		
PS8	1516025	19203	T	C	13	W5R
PS9	1516027	19221	A	G	31	T11A
PS10	1516029	19311	G	A		
PS11	1516039	21226	C	A	364	L122M
PS12 ^R	1516041	21250	T	C	388	C130R
PS13	1516045	21360	C	G	498	L166L
PS14 ^R	1516047	21388	C	T	526	R176C
PS15	1516051	21484	G	A	622	V208M
PS16	1516055	21818	C	T		

^aPolyId is a unique identifier assigned to each PS by Genaisance Pharmaceuticals, Inc.

^RPreviously reported in the literature

EXAMPLE 2

This example illustrates analysis of the APOE polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4. Genotypes and Haplotype Pairs Observed for APOE Gene																	
Genotype Number	Polymorphic Sites																HAP Pair
	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16	
1	T	C	G	C	G	G	C	T	A	G	C	T	C	C	G	C	1 1
2	T	C	G	G/C	G	G	C	T	A	G	C	T	C	C	G	C	1 2
3	T	C	G	G/C	G	G	C	T	A	G	C	T/C	C	C	G	C	1 3
4	T/C	C	G	G/C	G	G	C	T	A	G	C	T	C	C/T	G	C	1 4
5	T	C	G	G/C	G	G	C	T	A	G/A	C	T/C	C	C	G	C	1 5
6	T/C	C	G	G/C	G	G	C	T	A	G	C	T	C	C	G	C	1 6
7	T	C	G	G/C	G	G	C	T	A	G	C	T	C	C/T	G	C	1 7
8	T	C	G	G/C	G	G	C/A	T	A	G	C	T	C	C	G	C	1 9
9	T	C	G	C	G	G	C	T/C	A	G	C	T	C	C	G	C	1 11
10	T	C	G	C	G	G	C	T	A	G	C/A	T	C	C	G	C	1 14
11	T	C	G	G/C	G	G	C	T	A	G	C	T	C	C	G	C/T	1 17
12	T	C	G	G/C	G	G	C	T	A	G	C	T	C/G	-	G	C	1 18
13	T	C	G	G/C	G	G	C	T	A/G	G	C	T	C	C	G	C	1 19
14	T	C/T	G	G/C	G	G	C	T	A	G	C	T	C	C	G	C	1 20
15	T	C	G	G	G	G	C	T	A	G	C	T	C	C	G	C	2 2
16	T	C	G	G	G	G	C	T	A	G	C	T/C	C	C	G	C	2 3
17	T/C	C	G	G	G	G	C	T	A	G	C	T	C	C/T	G	C	2 4
18	T	C	G	G	G	G	C	T	A	G/A	C	T/C	C	C	G	C	2 5
19	T	C	G	G	G	G	C	T	A	G	C	T	C	C/T	G	C	2 7
20	T	C	G	G	G	G/A	C	T	A	G	C	T/C	C	C	G	C	2 8
21	T	C	G	G/C	G	G	C	T/C	A	G	C	T	C	C	G	C	2 11
22	T	C	G	G	G/A	G	C	T	A	G	C	T	C/G	C	G	C	2 12
23	T	C	G	G	G	G	C	T	A	G	C	T/C	C	C	G/A	C	2 13
24	T	C	G	G	G	G	C/A	T	A	G/A	C	T/C	C	C	G	C	2 15
25	T	C	G	G	G	G	C	T	A	G/A	C	T/C	C	C/T	G	C	2 16
26	-	C	G/A	G	G	G	C	T	A	G	C	T/C	C	C	G	C	2 21
27	T	C	G	G	G	G/A	C	T	A	G	C	C	C	C	G	C	3 8
28	T/C	C	G	G	G	G	C	T	A	G	C	C	C	C	G	C	3 10
29	C	C	G	G	G	G	C	T	A	G	C	T/C	C	C/T	G	C	4 10
30	C	C	G	G	G	G	C	T	A	G	C	T	C	C	G	C	6 6

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family)

and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 21 human APOE haplotypes shown in Table 5 below.

Table 5. Haplotypes Identified in the APOE Gene																
HAP No.	Polymorphic Sites															
	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16
1	T	C	G	C	G	G	C	T	A	G	C	T	C	C	G	C
2	T	C	G	G	G	G	C	T	A	G	C	T	C	C	G	C
3	T	C	G	G	G	G	C	T	A	G	C	C	C	C	G	C
4	C	C	G	G	G	G	C	T	A	G	C	T	C	T	G	C
5	T	C	G	G	G	G	C	T	A	A	C	C	C	C	G	C
6	C	C	G	G	G	G	C	T	A	G	C	T	C	C	G	C
7	T	C	G	G	G	G	C	T	A	G	C	T	C	T	G	C
8	T	C	G	G	G	A	C	T	A	G	C	C	C	C	G	C
9	T	C	G	G	G	G	A	T	A	G	C	T	C	C	G	C
10	C	C	G	G	G	G	C	T	A	G	C	C	C	C	G	C
11	T	C	G	C	G	G	C	C	A	G	C	T	C	C	G	C
12	T	C	G	G	A	G	C	T	A	G	C	T	G	C	G	C
13	T	C	G	G	G	G	C	T	A	G	C	C	C	C	A	C
14	T	C	G	C	G	G	C	T	A	G	A	T	C	C	G	C
15	T	C	G	G	G	G	A	T	A	A	C	C	C	C	G	C
16	T	C	G	G	G	G	C	T	A	A	C	C	C	T	G	C
17	T	C	G	G	G	G	C	T	A	G	C	T	C	C	G	T
18	T	C	G	G	G	G	C	T	A	G	C	T	G	C	G	C
19	T	C	G	G	G	G	C	T	G	G	C	T	C	C	G	C
20	T	T	G	G	G	G	C	T	A	G	C	T	C	C	G	C
21	T	C	A	G	G	G	C	T	A	G	C	C	C	C	G	C

In Table 6 below, the number of chromosomes in unrelated individuals characterized by a given haplotype is shown, arranged by ethnic background of the subjects in the Index Repository. In Table 7 below, the number of unrelated subjects characterized by a given haplotype is shown, again arranged by ethnic background of the subjects in the Index Repository. In Tables 6 and 7, the following abbreviations are used: AF, African or African-American; AS, Asian; CA, Caucasian; HL, Hispanic-Latino; and AM, Native Americans.

Table 6. Frequencies of Observed Haplotypes in Non-Related Individuals					
HAP No.	AF	CA	HL	AM	Total
1	8	20	16	4	66
2	17	13	13	1	55
3	5	1	1	0	7
4	1	2	0	0	6
5	0	2	2	0	5
6	1	2	0	0	5
7	0	0	2	0	2
8	2	0	0	0	2
9	1	0	1	0	2
10	2	0	0	0	2
11	0	0	0	0	2
12	0	1	0	0	1
13	1	0	0	0	1
14	0	0	0	0	1
15	1	0	0	0	1
16	0	1	0	0	1
17	0	0	0	1	1
18	1	0	0	0	1
19	0	0	1	0	1
20	0	0	0	0	1
21	0	0	0	0	1

Table 6. Frequencies of Observed Haplotype Pairs							
HAP Pair		AF	AS	CA	HL	AM	Total
1	1	1	3	3	3	1	11
2	1	2	7	9	5	1	24
2	2	5	0	0	3	0	8
3	1	1	0	1	0	0	2
3	2	2	0	0	1	0	3
4	1	0	2	2	0	0	4
4	2	0	1	0	0	0	1
5	1	0	0	0	2	0	2
5	2	0	1	2	0	0	3
6	1	1	0	2	0	0	3
6	6	0	1	0	0	0	1
7	1	0	0	0	1	0	1
7	2	0	0	0	1	0	1
8	2	1	0	0	0	0	1
8	3	1	0	0	0	0	1
9	1	1	0	0	1	0	2
10	3	1	0	0	0	0	1
10	4	1	0	0	0	0	1
11	1	0	1	0	0	0	1
11	2	0	1	0	0	0	1
12	2	0	0	1	0	0	1
13	2	1	0	0	0	0	1
14	1	0	1	0	0	0	1
15	2	1	0	0	0	0	1
16	2	0	0	1	0	0	1
17	1	0	0	0	0	1	1
18	1	1	0	0	0	0	1
19	1	0	0	0	1	0	1
20	1	0	1	0	0	0	1
21	2	0	1	0	0	0	1

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the apolipoprotein E (APOE) gene of an individual which comprises determining whether the individual has one of the APOE haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4.
2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS16 on at least one copy of the individual's APOE gene.
3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS16 on both copies of the individual's APOE gene.
4. A method for genotyping the apolipoprotein E (APOE) gene of an individual, comprising determining for the two copies of the APOE gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16.
5. The method of claim 4, wherein the determining step comprises:
 - (a) isolating from the individual a nucleic acid mixture comprising both copies of the APOE gene, or a fragment thereof, that are present in the individual;
 - (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
 - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
 - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
 - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
6. The method of claim 4, which comprises determining for the two copies of the APOE gene present in the individual the identity of the nucleotide pair at each of PS1-PS16.
7. A method for haplotyping the apolipoprotein E (APOE) gene of an individual which comprises determining, for one copy of the APOE gene present in the individual, the identity of the nucleotide at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16.
8. The method of claim 7, further comprising determining the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS12 and PS14.
9. The method of claim 7, wherein the determining step comprises:
 - (a) isolating from the individual a nucleic acid sample containing only one of the two copies

- of the APOE gene, or a fragment thereof, that is present in the individual;
- (b) amplifying from the nucleic acid molecule a target region containing the selected polymorphic site;
- (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
- (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
- (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
10. A method for predicting a haplotype pair for the apolipoprotein E (APOE) gene of an individual comprising:
- (a) identifying an APOE genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16;
- (b) enumerating all possible haplotype pairs which are consistent with the genotype;
- (c) comparing the possible haplotype pairs to the data in Table 4; and
- (d) assigning a haplotype pair to the individual that is consistent with the data.
11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS16.
12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the apolipoprotein E (APOE) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-22 shown in Table 5 and the haplotype pair is selected from the haplotype pairs shown in Table 4, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.
13. The method of claim 12, wherein the trait is a clinical response to a drug targeting APOE.
14. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the apolipoprotein E (APOE) gene at a polymorphic site selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16.
15. The composition of claim 14, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the APOE gene at a region containing

the polymorphic site.

16. The composition of claim 15, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-17, the complements of SEQ ID NOS:4-17, and SEQ ID NOS:18-45.
17. The composition of claim 14, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.
18. The composition of claim 17, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:46-73.
19. A kit for genotyping the APOE gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15 and PS16.
20. The kit of claim 19, which further comprises oligonucleotides designed to genotype each of PS12 and PS14.
21. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
 - (a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the apolipoprotein E (APOE) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises an APOE isogene defined by a haplotype selected from the group consisting of haplotypes 1-22 in Table 5; and
 - (b.) a second nucleotide sequence which is complementary to the first nucleotide sequence.
22. The isolated polynucleotide of claim 21, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
23. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 21, wherein the organism expresses an APOE protein encoded by the first nucleotide sequence.
24. The recombinant organism of claim 23, which is a nonhuman transgenic animal.
25. The isolated polynucleotide of claim 21, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the APOE gene, the fragment comprising one or more polymorphisms selected from the group consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, adenine at PS5, adenine at PS6, adenine at PS7, cytosine at PS8, guanine at PS9, adenine at PS10, adenine at PS11, guanine at PS13, adenine at PS15 and thymine at PS16.
26. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the APOE cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of an APOE isogene defined by one of the haplotypes shown in Table 5.

27. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 26, wherein the organism expresses a apolipoprotein E (APOE) protein encoded by the polymorphic variant sequence.
28. The recombinant organism of claim 27, which is a nonhuman transgenic animal.
29. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the APOE protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 5.
30. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 29.
31. A method for screening for drugs targeting the isolated polypeptide of claim 29 which comprises contacting the APOE polymorphic variant with a candidate agent and assaying for binding activity.
32. A computer system for storing and analyzing polymorphism data for the apolipoprotein E gene, comprising:
 - (a) a central processing unit (CPU);
 - (b) a communication interface;
 - (c) a display device;
 - 5 (d) an input device; and
 - (e) a database containing the polymorphism data;wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.
33. A genome anthology for the apolipoprotein E (APOE) gene which comprises APOE isogenes defined by any one of haplotypes 1-22 shown in Table 5.

1/7

POLYMORPHISMS IN THE APOE GENE

TTCATCTTAC	GTTAGTGGAA	ACTGAGGCTT	CCAGAGGTTT	CAAGGTCACA	
ACTAAATCCA	GAACCTCATC	TCAGGCACAC	TGGTCGTAGT	CCCAATGTCC	15200
AGTCTTAAGT	CTTCTTGGAT	ATCTGTGGCT	CACAGATTTT	GGGTGTTTGA	
GCCTCCTGCT	GAGCACTGCT	GGGGCCACAG	CGGTGACCAG	CCCTGTCTTC	15300
ACGGGACTCA	GTGAGAGGAA	CAGATTCATC	CGCAGAGTGG	GCAGGACTAG	
GTTGGGGGAA	CCCAGGGGTC	TAGAGGGCTT	TTCAGAGGGC	AGGGGTCACT	15400
GAGCGGAGAG	CAGAGGAGGA	GTGAGCCATT	TGCTCCAGCG	TGAAGTTGTT	
GGTGTGATGG	GGTTTCAGGG	TGGCAGGAGC	AGTGTGGTTA	AAGGTCTGGA	15500
AGCTGTCGGC	ATGTGGCTGG	TATCCAAGGT	GGCCAGGAAC	TCTGCATGGA	
TATGGTGGGA	AGCTGGCACG	CCTCTCACCT	CAGCTCTTCC	CTGCAGGCTC	15600
TGTGGATAGC	AACTGGATCG	TGGGTGCCAC	GCTGGAGAAG	AAGCTCCAC	
CCCTGCCCCCT	GACACTGGCC	CTTGGGGCCT	TCCTGAATCA	CCGCAAGAAC	15700
AAGTTTCAGT	GTGGCTTTGG	CCTCACCATC	GGCTGAGCCC	TCCTGGCCCC	
CGCCTTCCAC	GCCCTTCCGA	TTCCACCTCC	ACCTCCACCT	CCCCCTGCCA	15800
CAGAGGGGAG	ACCTGAGCCC	CCCTCCCTTC	CCTCCCCCCT	TGGGGGTCCG	
GGGGGACATT	GGAAAGGAGG	GACCCCGCCA	CCCCAGCAGC	TGAGGAGGGG	15900
ATTCTGGAAC	TGAATGGCGC	TTCGGGATTG	TGAGTAGCAG	GGGCAGCATG	
CCCAGTGGGC	CTGGGGTCCC	GGGAGGGATT	CCGGAATTGA	GGGGCACGCA	16000
GGATTCTGAG	CACCAAGGGC	AGAGGCGGCC	AGACAACCTC	AGGGAGGAGT	
GTCTTGGCGT	CCCCATCCTC	CAAAGGGCCT	GGGCCCCGCC	CGAGGGGGCA	16100
GCGAGAGGAG	CTTCCCCATC	CCCGTTCAGT	CCACCCTGCC	CCGTCCACTT	
TCCCATCTCC	TCGGTATAAA	TCATGTTTAT	AAGTTATGGA	AGAACC GGGA	16200
CATTTTACAG	AAAAAAAACA	AAAAACAACA	AAAAATATAC	GTGGGAAAAA	
AAACGATGGG	AGGCCCTCCG	TTTCTCAAGT	GTGTCTGGCC	TGTTTTGAGC	16300
ATTTTCATCCG	GAGTCTGGCC	GCCCTGACCT	TCCCCCAGCC	GCCTGCAGGG	
GGCGCCAGAG	GGCCGGAGCA	CGGAAAGCAG	CGGATCCTTG	ATGCTGCCTT	16400
AAGTCCGGCT	CAGAGGGGCG	CAGCGTGGCC	TGGGGTTCGCT	ATCTTCCCAT	
CCGGAACATC	TGCCCTGCTG	GGGGACACTA	CGGGCCTTCC	CTTGCCTGAG	16500
GGTAGGGTCT	CAAGGTCACT	TGCCCCCAGC	TTGACCTGGC	CGGAGTGGCT	
ATAGAGGACT	TTGTCCCTGC	AGACTGCAGC	AGCAGAGATG	ACACTGTCTC	16600
TGAGTGCAGA	GATGGGGGCA	GGGAGCTGGG	AGAGGGTTCA	AGCTACTGGA	
ACAGCTTCAG	AACAAC TAGG	GTACTAGGAA	CTGCTGTGTC	AGGGAGAAGG	16700
GGCTCAAGGA	CTCGCAGGCC	TGGGAGGAGG	GGCCTAGGCC	AGCCATGGGA	
GTTGGGTAC	CTGTGTCTGA	GGACTTGGTG	CTGTCTGGAT	TTTGCCAACC	16800
TAGGGCTGGG	GTCAGCTGAT	GCCCACCACG	ACTCCCGAGC	CTCCAGGAAC	
TGAAACCCTG	TCTGCCCCCA	GGGTCTGGGG	AAGGAGGCTG	CTGAGTAGAA	16900
CCAACCCAG	GTTACCAACC	CCACCTCAGC	CACCCCTTGC	CAGCCAAAGC	
AAACAGGCCC	GGCCCGGCAC	TGGGGGTTCC	TTCTCGAACC	AGGAGTTCAG	17000
CCTCCCCTGA	CCCGCAGAAT	CTTCTGATCC	CACCCGCTCC	AGGAGCCAGG	
AATGAGTCCC	AGTCTCTCCC	AGTTCTCACT	GTGTGGTTTT	GCCATTTCGTC	17100
TTGCTGCTGA	ACCACGGGTT	TCTCCTCTGA	AACATCTGGG	ATTTATAACA	
GGGCTTAGGA	AAGTGACAGC	GTCTGAGCGT	TACTGTGGC	CTGTCCATTG	17200
CTAGCCCTAA	CATAGGACCG	CTGTGTGCCA	GGGCTGTCCT	CCATGCTCAA	
TACACGTTAG	CTTGTCACCA	AACATACCCG	TGCCGCTGCT	TTCCCAGTCT	17300
GATGAGCAAA	GGAACCTGAT	GCTCAGAGAG	GACAAGTCAT	TTGCCCAAGG	
TCACACAGCT	GGCAACTGGC	AGAGCCAGGA	TTCACGCCCT	GGCAATTTGA	17400
CTCCAGAATC	CTAACCTTAA	CCCAGAAGCA	CGGCTTCAAG	CCCCTGGAAA	
CCACAATACC	TGTGGCAGCC	AGGGGGAGGT	GCTGGAATCT	CATTTACAT	17500
GTGGGGAGGG	GGCTCCCCTG	TGCTCAAGGT	CACAACCAAA	GAGGAAGCTG	
TGATTAATAAC	CCAGGTCCCA	TTTGCAAAGC	CTCGACTTTT	AGCAGGTGCA	17600
TCATACTGTT	CCCACCCCTC	CCATCCCACT	TCTGTCCAGC	CGCCTAGCCC	

FIGURE 1A

2/7					
CACTTTCTTT	TTTTTCTTTT	TTTGAGACAG	TCTCCCTCTT	GCTGAGGCTG	17700
GAGTGCAGTG	GCGAGATCTC	GGCTCACTGT	AACCTCCGCC	TCCCGGGTTC	
AAGCGATTCT	CCTGCCTCAG	CCTCCCAAGT	AGCTAGGATT	ACAGGCGCCC	17800
GCCACCACGC	CTGGCTAACT	TTTGTATTTT	TAGTAGAGAT	GGGGTTTCAC	
CATGTTGGCC	AGGCTGGTCT	CAAACCTCTG	ACCTTAAGTG	ATTGCCCCAC	17900
TGTGGCCTCC	CAAAGTGCTG	GGATTACAGG	CGTGACTACC	GCCCCCAGCC	
T					
CCTCCCATCC	CACTTCTGTC	CAGCCCCCTA	GCCCTACTTT	CTTTCTGGGA	18000
TCCAGGAGTC	CAGATCCCCA	GCCCCCTCTC	CAGATTACAT	TCATCCAGGC	
ACAGGAAAGG	ACAGGGTCAG	GAAAGGAGGA	CTCTGGGCGG	CAGCCTCCAC	18100
ATTCCCCTTC	CACGCTTGGC	CCCCAGAATG	GAGGAGGGTG	TCTGTATTAC	
TGGGCGAGGT	GTCCTCCCTT	CCTGGGGACT	GTGGGGGGTG	GTCAAAAGAC	18200
CTCTATGCCC	CACCTCCTTC	CTCCCTCTGC	CCTGCTGTGC	CTGGGGCAGG	
GGGAGAACAG	CCCACCTCGT	GA CTGGGGGC	TGGCCCAGCC	CGCCCTATCC	18300
CTGGGGGAGG	GGGCGGGACA	GGGGGAGCCC	TATAATTGGA	CAAGTCTGGG	
ATCCTTGAGT	CCTACTCAGC	CCCAGCGGAG	GTGAAGGACG	TCCTTCCCCA	18400
T					
GGAGCCGGTG	AGAAGCGCAG	TCGGGGGCAC	GGGGATGAGC	TCAGGGGCCT	
A					
CTAGAAAGAG	CTGGGACCCT	GGGAACCCCT	GGCCTCCAGG	TAGTCTCAGG	18500
G A					
AGAGCTACTC	GGGTCGGGC	TTGGGGAGAG	GAGGAGCGGG	GGTGAGGCAA	
A					
GCAGCAGGGG	ACTGGACCTG	GGAAGGGCTG	GGCAGCAGAG	ACGACCCGAC	18600
A					
CCGCTAGAAG	GTGGGGTG	GAGAGCAGCT	GGACTGGGAT	GTAAGCCATA	
GCAGGACTCC	ACGAGTTGTC	ACTATCATTT	ATCGAGCACC	TACTGGGTGT	18700
CCCCAGTGTC	CTCAGATCTC	CATAACTGGG	GAGCCAGGGG	CAGCGACACG	
GTAGCTAGCC	GTCGATTGGA	GAAC TTTAAA	ATGAGGACTG	AATTAGCTCA	18800
TAAATGGAAC	ACGGCGCTTA	ACTGTGAGGT	TGGAGCTTAG	AATGTGAAGG	
GAGAATGAGG	AATGCGAGAC	TGGGACTGAG	ATGGAACCGG	CGGTGGGGAG	18900
GGGGTGGGGG	GATGGAATTT	GAACCCCGGG	AGAGGAAGAT	GGAATTTTCT	
ATGGAGGCCG	ACCTGGGGAT	GGGGAGATAA	GAGAAGACCA	GGAGGGAGTT	19000
AAATAGGGAA	TGGGTTGGGG	GCGGCTTGGT	AAATGTGCTG	GGATTAGGCT	
GTTGCAGATA	ATGCAACAAG	GCTTGGAAGG	CTAACCTGGG	GTGAGGCCGG	19100
GTTGGGGCCG	GGCTGGGGGT	GGGAGGAGTC	CTCACTGGCG	GTTGATTGAC	
AGTTTCTCCT	TCCCCAGACT	GGCCAATCAC	AGGCAGGAAG	ATGAAGGTTC	19200
[exon 2: 19191..					
TGTGGGCTGC	GTTGCTGGTC	ACATTCCCTG	CAGGTATGGG	GGCGGGGCTT	
C G					
..19233]					
GCTCGGTTCC	CCCCGCTCCT	CCCCCTCTCA	TCCTCACCTC	AACCTCCTGG	19300
CCCCATTGAG	GCAGACCCTG	GGCCCCCTCT	TCTGAGGCTT	CTGTGCTGCT	
A					
TCCTGGCTCT	GAACAGCGAT	TTGACGCTCT	CTGGGCCTCG	GTTTCCCCCA	19400
TCCTTGAGAT	AGGAGTTAGA	AGTTGTTTTG	TTGTTGTTGT	TTGTTGTTGT	
TGTTTTGTTT	TTTTGAGATG	AAGTCTCGCT	CTGTCGCCCA	GGCTGGAGTG	19500
CAGTGGCGGG	ATCTCGGCTC	ACTGCAAGCT	CCGCCTCCCA	GGTCCACGCC	
ATTCTCCTGC	CTCAGCCTCC	CAAGTAGCTG	GGACTACAGG	CACATGCCAC	19600
CACACCCGAC	TAAC TTTTTT	GTATTTTCAG	TAGAGACGGG	GTTTCACCAT	
GTTGGCCAGG	CTGGTCTGGA	ACTCCTGACC	TCAGGTGATC	TGCCCGTTTC	19700
GATCTCCCAA	AGTGCTGGGA	TTACAGGCGT	GAGCCACCGC	ACCTGGCTGG	

FIGURE 1B

3/7

GAGTTAGAGG TTTCTAATGC ATTGCAGGCA GATAGTGAAT ACCAGACACG 19800
 GGGCAGCTGT GATCTTTTATT CTCCATCACC CCCACACAGC CCTGCCTGGG
 GCACACAAGG ACACTCAATA CATGCTTTTC CGCTGGGCGC GGTGGCTCAC 19900
 CCCTGTAATC CCAGCACTTT GGGAGGCCAA GGTGGGAGGA TCACTTGAGC
 CCAGGAGTTC AACACCAGCC TGGGCAACAT AGTGAGACCC TGTCTCTACT 20000
 AAAAATACAA AAATTAGCCA GGCATGGTGC CACACACCTG TGCTCTCAGC
 TACTCAGGAG GCTGAGGCAG GAGGATCGCT TGAGCCCAGA AGGTCAAGGT 20100
 TGCAGTGAAC CATGTTTCAGG CCGCTGCACT CCAGCCTGGG TGACAGAGCA
 AGACCCCTGTT TATAAATACA TAATGCTTTC CAAGTGATTA AACCGACTCC 20200
 CCCCTCACCC TGCCCACCAT GGCTCCAAAG AAGCATTTGT GGAGCACCTT
 CTGTGTGCCC CTAGGTACTA GATGCCTGGA CGGGGTGAGA AGGACCCTGA 20300
 CCCACCTTGA ACTTGTTCCA CACAGGATGC CAGGCCAAGG TGGAGCAAGC
 [exon 3: 20326..
 GGTGGAGACA GAGCCGGAGC CCGAGCTGCG CCAGCAGACC GAGTGGCAGA 20400
 GCGGCCAGCG CTGGGAACTG GCACTGGGTC GCTTTTGGA TTACCTGCGC
 TGGGTGCAGA CACTGTCTGA GCAGGTGCAG GAGGAGCTGC TCAGCTCCCA 20500
 GGTCACCCAG GAACTGAGGT GAGTGTCCCC ATCCTGGCCC TTGACCCTCC
 ..20518]
 TGGTGGGCGG CTATACCTCC CCAGGTCCAG GTTTCATTCT GCCCCTGTCTG 20600
 CTAAGTCTTG GGGGGCCTGG GTCTCTGCTG GTTCTAGCTT CCTCTTCCCA
 TTTCTGACTC CTGGCTTTAG CTCTCTGGAA TTCTCTCTCT CAGCTTTGTC 20700
 TCTCTCTCTT CCCTTCTGAC TCAGTCTCTC ACACCTCGTC TGGCTCTGTC
 TCTGTCTCTT CCTAGCTCTT TTATATAGAG ACAGAGAGAT GGGGTCTCAC 20800
 TGTGTGCCC AGGCTGGTCT TGAACCTCTG GGCTCAAGCG ATCCTCCCGC
 CTCGGCCTCC CAAAGTGCTG GGATTAGAGG CATGAGCCAC CTGCCCCGGC 20900
 CTCTAGCTC CTCTTCTGTC TCTGCCTCTG CCCTCTGCAT CTGCTCTCTG
 CATCTGTCTC TGTCTCCTTC TCTCGGCCTC TGCCCCGTTT CTCTCTCTCC 21000
 TCTTGGGTCT CTCTGGCTCA TCCCCATCTC GCCCGCCCCA TCCCAGCCCT
 TCTCCCCGCC TCCCACTGTG CGACACCCTC CCGCCCTCTC GGCCGCAAGG 21100
 [exon 4: 21099..
 CGCTGATGGA CGAGACCATG AAGGAGTTGA AGGCCTACAA ATCGGAACTG
 GAGGAACAAC TGACCCCGGT GCGGAGGAG ACGCGGGCAC GGCTGTCCAA 21200
 GGAGCTGCAG GCGGCGCAGG CCCGGCTGGG CGCGGACATG GAGGACGTGT
 A C
 GCGGCCGCCT GGTGCAGTAC CGCGGCGAGG TGCAGGCCAT GCTCGGCCAG 21300
 AGCACCGAGG AGCTGCGGGT GCGCCTCGCC TCCCACCTGC GCAAGCTGCG
 TAAGCGGCTC CTCCGCGATG CCGATGACCT GCAGAAGCGC CTGGCAGTGT 21400
 G T
 ACCAGGCCGG GGGCCGCGAG GCGCGCGAGC GCGGCCTCAG CGCCATCCGC
 GAGCGCCTGG GGGCCCTGGT GGAACAGGGC CGCGTGCGGG CCGCCACTGT 21500
 A
 GGGCTCCCTG GCCGGCCAGC CGCTACAGGA GCGGGCCCAG GCCTGGGGCG
 AGCGGCTGCG CGCGCGGATG GAGGAGATGG GCAGCCGGAC CCGCGACCGC 21600
 CTGGACGAGG TGAAGGAGCA GGTGGCGGAG GTGCGCGCCA AGCTGGAGGA
 GCAGGCCAG CAGATACGCC TGCAGGCCGA GGCCTTCCAG GCCCGCCTCA 21700
 AGAGCTGGTT CGAGCCCCTG GTGGAAGACA TGCAGCGCCA GTGGGCGGGG
 CTGGTGGAGA AGGTGCAGGC TGCCGTGGGC ACCAGCGCCG CCCCTGTGCC 21800
 CAGCGACAAT CACTGAACGC CGAAGCCTGC AGCCATGCGA CCCACGCCA
 T
 ..21816]
 CCCCCTGCCT CCTGCCTCCG CGCAGCCTGC AGCGGGAGAC CCTGTCCCCG 21900
 CCCAGCCGT CCTCCTGGGG TGGACCCTAG TTTAATAAAG ATTCACCAAG

FIGURE 1C

4/7

TTTCACGCAT	CTGCTGGCCT	CCCCCTGTGA	TTTCCTCTAA	GCCCCAGCCT	22000
CAGTTTCTCT	TTCTGCCCCAC	ATACTGGCCA	CACAATTCTC	AGCCCCCTCC	
TCTCCATCTG	TGTCTGTGTG	TATCTTTCTC	TCTGCCCTTT	TTTTTTTTTT	22100
TAGACGGAGT	CTGGCTCTGT	CACCCAGGCT	AGAGTGCAGT	GGCACGATCT	
TGGCTCACTG	CAACCTCTGC	CTCTTGGGTT	CAAGCGATTG	TGCTGCCTCA	22200
GTAGCTGGGA	TTACAGGCTC	ACACCACCAC	ACCCGGCTAA	TTTTTGTATT	
TTTAGTAGAG	ACGAGCTTTC	ACCATGTTGG	CCAGGCAGGT	CTCAAACCTC	22300
TGACCAAGTG	ATCCACCCGC	CGGCCTCCCA	AAGTGCTGAG	ATTACAGGCC	
TGAGCCACCA	TGCCCGGCCT	CTGCCCCCTC	TTCTTTTTTA	GGGGGCAGGG	22400
AAAGGTCTCA	CCCTGTCACC	CGCCATCACA	GCTCACTGCA	GCCTCCACCT	
CCTGGACTCA	AGTGATAAGT	GATCCTCCCG	CCTCAGCCTT	TCCAGTAGCT	22500
GAGACTACAG	GCGCATACCA	CTAGGATTAA	TTTGGGGGGG	GGGTGGTGTG	
TGTGGAGATG	GGGTCTGGCT	TTGTTGGCCA	GGCTGATGTG	GAATTCCTGG	22600
GCTCAAGCGA	TACTCCCAAC	TTGGCCTCCT	GAGTAGCTGA	GACTACTGGC	
TAGCACCACC	ACACCCAGCT	TTTTATTATT	ATTGTAGAG	ACAAGGTCTC	22700
AATATGTTGC	CCAGGCTAGT	CTCAAACCCC	TGGGCTCAAG	AGATCCTCCG	
CCATCGGCCT	CCCAAAGTGC	TGGGATTCCA	GGCATGGGGC	TCCGAGCCCG	22800
GCCTGCCCAA	CTTAATAATA	CTTGTTCCCTC	AGAGTTGCAA	CTCCAAATGA	
CCTGAGATTG	GTGCCTTTAT	TCTAAGCTAT	TTTCATTTTT	TTTCTGCTGT	22900
CATTATTCTC	CCCCTTCTCT	CCTCCAGTCT	TATCTGATAT	CTGCCTCCTT	
CCCACCCACC	CTGCACCCCA	TCCCACCCCT	CTGTCTCTCC	CTGTTCTCCT	23000
CAGGAGACTC	TGGCTTCCTG	TTTTCCCTCCA	CTTCTATCTT	TTATCTCTCC	
CTCCTACGGT	TTCTTTTCTT	TCTCCCCGGC	CTGCTTGTTT	CTCCCCAAC	23100
CCCCTTCATC	TGGATTTCTT	CTTCTGCCAT	TCAGTTTGGT	TTGAGCTCTC	
TGCTTCTCCG	GTTCCCTCTG	AGCTAGCTGT	CCCTTCACCC	ACTGTGAAC	23200
GGGTTTCCCT	GCCCAACCCCT	CATTCTCTTT	CTTCTTTTCT	TTTTTTTTTT	
TTTTTTTTTT	TTTTTTTTTT	GAGACAGAGT	CTTGCTCTGT	TGCCCAGCCT	23300
GGAGTGCAGT	GGTGCAATCT	TGGTTCCTG	CAACCTCCAC	TTCCCAGATT	
CAAGCAATTC	TCCTGCCTCA	GCCTCCAGAG	TAGCTGGGAT	TACAGGCGTG	23400
TCCCACCACA	CCCGACTAAT	TTTTGTATTT	TTGGTAGAGA	CAAGGCTTCG	
GCATTGTTGG	CCAGGCAGGT	CTCGAACTCC	TGACCTCAAG	TAATCTGCCT	23500
GCCTCACCCCT	CCCAAAGTGC	TGGGATTACA	GGCATGAGCC	ACCTCACCCG	
GACCATCCCT	CATTCTCCAT	CCTTTCCTCC	AGTTGTGATG	TCTACCCCTC	23600
ATGTTTCCCA	ACAAGCCTAC	TGGGTGCTGA	ATCCAGGCTG	GGAAGAGAAG	
GGAGCGGCTC	TTCTGTCGGA	GTCTGCACCA	GGCCCATGCT	GAGACGAGAG	23700
CTGGCGCTCA	GAGAGGGGAA	GCTTGGATGG	AAGCCCAGGA	GCCGCCGGCA	
CTCTCTTCTC	CTCCCACCCC	CTCAGTTCTC	AGAGACGGGG	AGGAGGGTTC	23800
CCACCAACGG	GGGACAGGCT	GAGACTTGAG	CTTGTATCTC	CTGGGCCAGC	
TGCAACATCT	GCTTGTCCCT	CTGCCCATCT	TGGCTCCTGC	ACACCCTGAA	23900
CTTGGTGCTT	TCCCTGGCAC	TGCTCTGATC	ACCCACGTGG	AGGCAGCACC	
CCTCCCCTGG	AGATGACTCA	CCAGGGCTGA	GTGAGGAGGG	GAAGGGTCAG	24000
TGTGCTCACA	GGCAGGGGGC	CTGGTCTGCT	GGGCCTGCTG	CTGATTCAAC	
GTATGTCCAG	GAGCTGGAGC	AGCACCCCTC	CAGAGCCTCA	GTTTCCCCAG	24100
ATGTCAATGA	GAGAATCAGC	ATCAGCCACA	TCTCCACCT	GAAGAATCGA	
ACCTTGAGTT	CTCACCTTTA	AGAATTTTTT	TTTTTTTTTG	AGACAGTGCC	24200
TTGCTEAGTC	GACCACGCTG	GAGTGCAGTA	GCACGATCAT	CAATCACTGC	
AGCCTCCAAC	TCAAGCAACC	CTCCTGCCTC	AACCTCCTGA	GTAGCTGGGG	24300
CCACAGGCAC	CCACCACCAT	GCCTGGCTAA	TTCTTTCTTT	TTTGAGATGG	
AGTCTCACTC	TGTCGCCCCAG	GCTGGAGTGC	AGTGGCACCA	TCTTGGCTCA	24400
CCGCAACCTC	TGCCCCCCGG	GTTCAAGCAA	TTCTCCTGCC	TCAGCCTCCT	
GAGTAGCTAG	AATTACAGGC	GCACACCACC	ACACCAGGCT	AATTTTTGTA	24500
TTTTTAGTAG	AAACGAGGTT	TCACCACGGT	GGTCAGGTTG	GTCTCAAAC	

FIGURE 1D

5/7					
CCTGACATCA	GGTGATCCAC	CCGCCTCGGC	CTCCCAAAGC	ACTGGGATTA	24600
CAGGAGTGAG	CCACCATGCC	TGGCCATATT	TTAAACTTTT	TTGTAGAGAT	
GGGGTCTTAC	TATGTTGCCC	AGGCTGATCT	CAAACCTCTG	GGCTCAAGCA	24700
GATCCTCCTG	CCAAAGCTTC	CCAAAGTAGT	TGGATTACAG	GTGTGAGCCA	
CCACGCCCCA	CCAACCTATG	AAGATGTCTA	AGCAGACAGG	GTGCCCTATC	24800
TCACCACATC	CCCAGTCAGA	TCCAGAGGTG	GGACAAGGAT	GGGAAGGAGG	
TTCCAGAACA	AAGGCTGGCA	GGGAGTCGTA	TGGGACAGGA	GCTGACCCAG	24900
CAACCATCCA	CAGAGACATC	CTGGAGCCTG	GGAAGGAGAA	GGACAAAGAG	
CCCCCTTTTT	TAAATTTTTT	TTATGTTTTT	GAGACGGAGT	CTCACTCTGT	25000
CACCCAAGCT	GGAGTGCAGT	GGCACAATCT	TGGCTCACTG	CAACCTCCAC	
CTCATGGGTT	CAAACGACTC	TCCTGCCTCA	GCCTCCTGAG	TAGCTGGGAC	25100
CACAGGTGCA	CACCACCATG	CCTGGATAAT	TTTTGAATTT	TTGGTAGAGA	
CGGAGTTTCA	CCATGTTGGC	CAGGCAGGTC	TCGAACTCCT	GACCTCAAGT	25200
GCTCCGCCCA	CCTTGGCCTC	CCAAAGTTCT	GGGAATACAG	GCGTGAGCCA	
CTGCAACCAG	CCAGTAGCCC	CCATCTTTGC	CCCTCGCTGA	GCCCTACTGG	25300
ATGTTCTTGG	TTATGCGACA	GTTTCCCCAT	CTATTAAAGA	GAAACCCCTA	
TAGCAGAGGG	GAGGATGAGG	TTGGAAAAGC	AGGAGCATTG	TTATGCTATT	25400
CTTGTGGGGT	CTGGGAAGCA	GACATCTGGG	TGGATGTTTG	GGGGGTGCTG	
GGCTTAGTTG	GGGAAGTAGG	GGGGCCCCTG	GGGCTGACAG	GGAAGGGAAG	25500
CTCTGAGCTG	GCCAGAGGGA	TGTTGCAATC	CTGCCAGGGT	CTTGTCTATG	
CTGTCCTTTT	CACAACCATC	CCCCTACTGC	CAGGCTGACA	CGTGGTTGCG	25600
GGGGCACAAG	GCCAGCCAAC	CTAGAGTCTG	AGGCTAGGCG	GAGGACACCC	
TCCCCACCAG	CTGCCAGGGT	CACTGGCGGT	CAAAGGCAGC	TGGTGGGGAA	25700
GGCATTGGAC	TCCAGCCTTG	GGGGACGGAT	GTAGTGATGG	TGGGAAGCAG	
GCTTGGTGCC	AGGAGGGGCG	TCAGAGGGTG	AATAAAAGCA	GATAGAGTGT	25800
TTGGGGGAGG	TAGCCA				25816

FIGURE 1E

6/7

POLYMORPHISMS IN THE CODING SEQUENCE OF APOE

ATGAAGGTTT	TGTGGGCTGC	GTTGCTGGTC	ACATTCCTGG	CAGGATGCCA	
	C		G		
GGCCAAGGTG	GAGCAAGCGG	TGGAGACAGA	GCCGGAGCCC	GAGCTGCGCC	100
AGCAGACCGA	GTGGCAGAGC	GGCCAGCGCT	GGGAAGTGGC	ACTGGGTCGC	
TTTTGGGATT	ACCTGCGCTG	GGTGCAGACA	CTGTCTGAGC	AGGTGCAGGA	200
GGAGCTGCTC	AGCTCCCAGG	TCACCCAGGA	ACTGAGGGCG	CTGATGGACG	
AGACCATGAA	GGAGTTGAAG	GCCTACAAAT	CGGAAGTGGG	GGAACAAGTG	300
ACCCCGGTGG	CGGAGGAGAC	GCGGGCACGG	CTGTCCAAGG	AGCTGCAGGC	
GGCGCAGGCC	CGGCTGGGCG	CGGACATGGA	GGACGTGTGC	GGCCGCCTGG	400
	A		C		
TGCAGTACCG	CGGCGAGGTG	CAGGCCATGC	TCGGCCAGAG	CACCGAGGAG	
CTGCGGGTGC	GCCTCGCCTC	CCACCTGCGC	AAGCTGCGTA	AGCGGCTCCT	500
			G		
CCGCGATGCC	GATGACCTGC	AGAAGCGCCT	GGCAGTGTAC	CAGGCCGGGG	
		T			
CCCGCGAGGG	CGCCGAGCGC	GGCCTCAGCG	CCATCCGCGA	GCGCCTGGGG	600
CCCCTGGTGG	AACAGGGCCG	CGTGCGGGCC	GCCACTGTGG	GCTCCCTGGC	
	A				
CGGCCAGCCG	CTACAGGAGC	GGGCCCAGGC	CTGGGGCGAG	CGGCTGCGCG	700
CGCGGATGGA	GGAGATGGGC	AGCCGGACCC	GCGACCGCCT	GGACGAGGTG	
AAGGAGCAGG	TGGCGGAGGT	GCGCGCCAAG	CTGGAGGAGC	AGGCCAGCA	800
GATACGCCTG	CAGGCCGAGG	CCTTCCAGGC	CCGCCTCAAG	AGCTGGTTTCG	
AGCCCCTGGT	GGAAGACATG	CAGCGCCAGT	GGGCCGGGCT	GGTGGAGAAG	900
GTGCAGGCTG	CCGTGGGCAC	CAGCGCCGCC	CCTGTGCCCA	GCGACAATCA	
CTGA					954

FIGURE 2

7/7

ISOFORMS OF THE APOE PROTEIN

MKVLWAALLV	TFLAGCQAKV	EQAVETEPEP	ELRQQTEWQS	QQRWELALGR	
R	A				
FWDYLRWVQT	LSEQVQEELL	SSQVTQELRA	LMDETMKELK	AYKSELEEQL	100
TPVAEETRAR	LSKELQAAQA	RLGADMEDVC	GRLVQYRGEV	QAMLGQSTEE	
		M	R		
LRVRLASHLR	KLRKRLLRDA	DDLQKRLAVY	QAGAREGAER	GLSAIRERLG	200
		C			
PLVEQGRVRA	ATVGSLAGQP	LQERAQAWGE	RLRARMEEMG	SRTRDRLDEV	
M					
KEQVAEVRAR	LEEQAQQIRL	QAEAFQARLK	SWFEPLVEDM	QRQWAGLVEK	300
VQAAVGTSAA	PVPSDNH				317

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.

Choi, Julie Y.

Kliem, Stefanie E.

Koshy, Beena

Lee, Helen H.

<120> HAPLOTYPES OF THE APOE GENE

<130> MWH-0371PCT APOE

<140> tba

<141> 2001-04-16

<150> 60/197,188

<151> 2000-04-14

<160> 76

<170> PatentIn Ver. 2.1

<210> 1

<211> 10716

<212> DNA

<213> Homo sapiens

<400> 1

```

ttcatcttac gttagtggaa actgaggctt ccagaggttt caaggtcaca actaaatcca 60
gaacctcatc tcaggcacac tggtcgtagt cccaatgtcc agtcttaagt cttcttggat 120
atctgtggct cacagatttt ggggtgttga gcctcctgct gagcactgct ggggccacag 180
cggtgaccag cctgtctctt acgggactca gtgagaggaa cagattcatc cgcagagtgg 240
gcaggactag gttgggggaa cccaggggtc tagagggctt ttcagagggc aggggtcact 300
gagcggagag cagaggagga gtgagccatt tgctccagcg tgaagttgtt ggtgtgatgg 360
ggttttcagg tggcaggagc agtgtggtta aaggtctgga agctgtcggc atgtggctgg 420
tatccaaggt ggccaggaac tctgcatgga tatggtggga agctggcacg cctctcacct 480
cagctcttcc ctgcaggctc tgtggatagc aactggatcg tgggtgccac gctggagaag 540
aagctccac cctgccccct gacactggcc cttggggcct tctgaatca ccgcaagaac 600
aagtttcagt gtggttttg cctcaccatc ggctgagccc tctggcccc cgccttccac 660
gcccttccga ttccacctcc acctccacct cccctgccca cagaggggag acctgagccc 720
ccctcccttc cctccccctc tgggggtcgg gggggacatt ggaaaggagg gaccccgcca 780
ccccagcagc tgaggagggg attctggaac tgaatggcgc ttcgggattc tgagtagcag 840
gggcagcatg cccagtgggc ctgggggtccc gggagggatt ccggaattga ggggcacgca 900
ggattctgag caccaggggc agaggcggcc agacaacctc agggaggagt gtcctggcgt 960
ccccatcctc caaagggcct gggcccgcgc cgagggggca gcgagaggag cttccccatc 1020
cccggtcagt ccacctgcc ccgtccactt tcccatctcc tcggtataaa tcatgtttat 1080
aagttatgga agaaccggga cattttacag aaaaaaaca aaaaacaaca aaaaatatac 1140
gtgggaaaaa aaacgatggg aggcctccgt tttctcaagt gtgtctggcc tgttttgagc 1200

```

atttcatccg gagtctggcc gccctgacct tccccagcc gcctgcaggg ggcgccagag 1260
 ggccggagca cggaaagcag cggatccttg atgctgcctt aagtcgggct cagaggggcg 1320
 cagcgtggcc tggggctcgt atcttcccat ccggaacatc tgccctgctg ggggacacta 1380
 cgggccttcc cttgcctgag ggtaggggtct caaggctact tgccccagc ttgacctggc 1440
 cggagtggct atagaggact ttgtccctgc agactgcagc agcagagatg acactgtctc 1500
 tgagtgcaga gatgggggca gggagctggg agaggggttca agctactgga acagcttcag 1560
 aacaactagg gtactaggaa ctgctgtgtc agggagaagg ggctcaagga ctgcaggcc 1620
 tgggaggagg ggcctaggcc agccatggga gttgggtcac ctgtgtctga ggacttgggtg 1680
 ctgtctggat ttgtccaaac tagggctggg gtcagctgat gcccaccacg actcccagc 1740
 ctccaggaac tgaaaccttg tctgccccca gggctctggg aaggaggctg ctgagtagaa 1800
 ccaacccag gttaccaacc ccacctcagc cacccttgc cagccaaagc aaacaggccc 1860
 ggccggcac tgggggttcc ttctcgaacc aggagtccag cctcccctga cccgcagaat 1920
 cttctgatcc caccgctcc aggagccagg aatgagctcc agtctctccc agttctcact 1980
 gtgtggtttt gccattcgct ttgctgtgta accacgggtt tctcctctga aacatctggg 2040
 atttataaca gggcttagga aagtgcagc gtctgagcgt tcaactgtggc ctgtccattg 2100
 ctagccctaa cataggaccg ctgtgtgcca gggctgtcct ccatgctcaa tacacgttag 2160
 cttgtcacca aacatacccg tgccgctgct tcccagctct gatgagcaaa ggaacttgat 2220
 gctcagagag gacaagtcac ttgcccagg tcacacagct ggcaactggc agagccagga 2280
 ttcacgcctt ggcaatttga ctccagaatc ctaaccttaa cccagaagca cggcttcaag 2340
 cccctggaaa ccacaatacc tgtggcagcc agggggagggt gctggaatct catttcacat 2400
 gtggggaggg ggctcccctg tgctcaaggt cacaaccaa gaggaagctg tgattaaaac 2460
 ccaggctcca ttgtcaaagc ctgcactttt agcagggtgca tcatactgtt cccaccctc 2520
 ccatcccact tctgtccagc cgcctagccc cactttcttt tttttctttt tttagacag 2580
 tctccctctt gctgaggctg gagtgcagtg gcgagatctc ggctcactgt aacctccgcc 2640
 tcccggttc aagcgattct cctgcctcag cctcccaagt agctaggatt acaggcgccc 2700
 gccaccacgc ctggctaact ttgtatttt tagtagagat ggggtttcac catgttggcc 2760
 aggctggtct caaactcctg accttaagtg attcgcccac tgtggcctcc caaagtgtg 2820
 ggattacagg cgtgactacc gccccagcc cctcccatcc cacttctgtc cagcccccta 2880
 gccctacttt ctttctggga tccaggagtc cagatcccca gcccctctc cagattacat 2940
 tcatccaggc acaggaaagg acagggtcag gaaaggagga ctctgggcgg cagcctccac 3000
 attcccttcc cacgcttggc cccagaatg gaggagggtg tctgtattac tgggcgaggt 3060
 gtctccctt cctggggact gtggggggtg gtcaaaagac ctctatgcc cacctcttc 3120
 ctccctctgc cctgctgtgc ctggggcagg gggagaacag cccacctcgt gactgggggc 3180
 tggcccagcc cgcctatcc ctgggggagg gggcgggaca gggggagccc tataattgga 3240
 caagtctggg atccttgagt cctactcagc cccagcggag gtgaaggacg tccttcccca 3300
 ggagccgggt agaagcgag tcgggggcac ggggatgagc tcaggggcct ctagaaagag 3360
 ctgggacctt gggaacccct ggcctccagg tagtctcagg agagctactc ggggtcgggc 3420
 ttggggagag gaggagcggg ggtgaggcaa gcagcagggg actggacctg ggaagggtg 3480
 ggcagcagag acgaccgac ccgctagaag gtgggggtgg gagagcagct ggactgggat 3540
 gtaagccata gcaggactcc acgagttgtc actatcattt atcgagcacc tactgggtgt 3600
 cccagtgctc ctcatgctc cataactggg gagccagggg cagcgacacg gtagctagcc 3660
 gtcgattgga gaactttaa atgaggactg aattagctca taaatggaac acggcgctta 3720
 actgtgaggt tggagcttag aatgtgaagg gagaatgagg aatgcgagac tgggactgag 3780
 atggaaccgg cgggtggggg ggggtggggg gatggaattt gaaccccggg agaggaagat 3840
 ggaattttct atggaggccg acctggggat ggggagataa gagaagacca ggaggaggt 3900
 aaatagggaa tgggttgggg gcggcttggg aaatgtgctg ggattaggct gttgcagata 3960
 atgcaacaag gcttggaagg ctaacctggg gtgaggccgg gttggggccg ggctgggggt 4020
 gggaggagtc ctactggcg gttgattgac agtttctcct tcccagact ggccaatcac 4080

aggcaggaag atgaagggtc tgtgggctgc gttgctggtc acattcctgg caggatatggg 4140
 ggcgggggctt gctcgggtcc ccccgctcct cccctctca tcctcacctc aacctcctgg 4200
 ccccatcag gcagaccctg ggccccctct tctgaggctt ctgtgctgct tcctggctct 4260
 gaacagcgat ttgacgctct ctgggcctcg gtttcccca tccttgagat aggagttaga 4320
 agttgttttg ttgttggtgt ttgttggtgt tgttttggtt ttttgagatg aagtctcgct 4380
 ctgtcgccca ggctggagtg cagtggcggg atctcggtc actgcaagct ccgcctccca 4440
 ggtccacgcc attctcctgc ctcagcctcc caagtagctg ggactacagg cacatgccac 4500
 cacacccgac taactttttt gtattttcag tagagacggg gtttcacat gttggccagg 4560
 ctggtctgga actcctgacc tcagggtgat tgcctgttc gatctccaa agtgctggga 4620
 ttacaggcgt gagccaccgc acctggctgg gagttagagg tttctaagtc attgcaggca 4680
 gatagtgaat accagacacg gggcagctgt gatctttatt ctccatcacc cccacacagc 4740
 cctgcctggg gcacacaagg aactcaata catgcttttc cgctgggcgc ggtggctcac 4800
 cctgtaate ccagcacttt gggaggccaa ggtgggagga tcacttgagc ccaggagttc 4860
 aacaccagcc tgggcaacat agtgagacc tgtctctact aaaaatacaa aaattagcca 4920
 ggcattggtc cacacacctg tgctctcagc tactcaggag gctgaggcag gaggatcgct 4980
 tgagcccaga aggtcaaggt tgcagtgaac catgttcagg ccgctgcact ccagcctggg 5040
 tgacagagca agaccctgtt tataaatata taatgctttc caagtgatta aaccgactcc 5100
 cccctcacc tgcccaccat ggctccaaag aagcatttgt ggagcacctt ctgtgtgccc 5160
 ctaggtacta gatgcctgga cggggtcaga aggacctga cccacctga acttgttcca 5220
 cacaggatgc caggccaagg tggagcaagc ggtggagaca gagccggagc ccgagctgcg 5280
 ccagcagacc gagtggcaga gcggccagcg ctgggaactg gcaactgggtc gcttttggga 5340
 ttacctgcgc tgggtgcaga cactgtctga gcagggtcag gaggagctgc tcagctccca 5400
 ggtcaccag gaactgaggt gagtgtcccc atcctggccc ttgacctcc tggtgggcgg 5460
 ctatacctcc ccaggctcag gtttcattct gccctgtcg ctaagtcttg gggggcctgg 5520
 gtctctgctg gttctagctt cctcttccca tttctgactc ctggctttag ctctctggaa 5580
 ttctctctct cagctttgtc tctctctctt ccctctgac tcagtctctc acactcgtcc 5640
 tggtctgtc tctgtccttc cctagctctt ttatatagag acagagagat ggggtctcac 5700
 tgtgttgccc aggttggtct tgaacttctg ggctcaagcg atcctccgc ctcggcctcc 5760
 caaagtgtct ggattagagg catgagccac cttgcccggc ctctagctc cttcttcgtc 5820
 tctgcctctg ccctctgcat ctgctctctg catctgtctc tgtctccttc tctcggctc 5880
 tgccccgttc cttctctccc tcttggtct ctctggctca tcccatctc gccgcacca 5940
 tcccagcctt tctccccgc tccactgtg cgacaccctc ccgcctctc ggccgcaggg 6000
 cgctgatgga cgagaccatg aaggagttga aggcctacaa atcggaactg gaggaacaac 6060
 tgacccccgt ggcggaggag acgcgggcac ggctgtccaa ggagctgcag gcggcgagg 6120
 cccggtggg cgcggacatg gaggacgtgt gcggccgctt ggtgcagtag cgcggcgagg 6180
 tgcaggccat gctcggccag agcaccgagg agctgcgggt gcgcctcgcc tcccacctgc 6240
 gcaagctgcg taagcggctc ctccgcgat ccgatgacct gcagaagcgc ctggcagtg 6300
 accaggccgg ggcgcgcgag ggcgcgcgag gcggcctcag cgcctccgc gagcgcctgg 6360
 ggccccgtt ggaacagggc cgcgtgcggg ccgccaactg gggctccctg gccggccagc 6420
 cgctacagga gcgggcccag gctggggcg agcggtgcg cgcgcggatg gaggagatgg 6480
 gcagccggac ccgcgaccgc ctggacgagg tgaaggagca ggtggcggag gtgcgcgcca 6540
 agctggagga gcaggcccag cagatacgcc tgcaggccga ggccttcag gccgcctca 6600
 agagctggtt cgagccccgt gtggaagaca tgcagcgcca gtgggcccgg ctggtggaga 6660
 aggtgcaggc tgccgtgggc accagcgccg cccctgtgcc cagcgacaat cactgaacgc 6720
 cgaagcctgc agccatgcga cccacgcca cccgtgcct cctgcctccg cgcagcctgc 6780
 agcgggagac cctgtccccg cccagccgt cctcctgggg tggaccctag ttaataaag 6840
 attcaccaag tttcacgcat ctgctggcct cccctgtga tttcctctaa gcccagcct 6900
 cagtttctct ttctgcccac ataactggcca cacaattctc agccccctcc tctccatctg 6960

tgtctgtgtg tatctttctc tctgcccttt tttttttttt tagacggagt ctggctctgt 7020
 caccaggt agagtgcagt ggcacgatct tggctcactg caacctctgc ctcttgggtt 7080
 caagcgattc tgctgcctca gtagctggga ttacaggctc acaccaccac acccggttaa 7140
 tttttgtatt tttagtagag acgagctttc accatgttgg ccaggcaggt ctcaaactcc 7200
 tgaccaagt atccaccgc cgccctccca aagtgtgag attacaggcc tgagccacca 7260
 tgcccgccct ctgccctct tcttttttta gggggcagg aaaggtctca ccctgtcacc 7320
 cgccatcaca gctcactgca gcctccacct cctggactca agtgataagt gatcctcccg 7380
 cctcagcctt tccagtagct gagactacag gcgcatacca ctaggattaa tttggggggg 7440
 ggggtgtgtg tgtggagatg gggctgtggt ttgttggcca ggctgatgtg gaattcctgg 7500
 gctcaagcga tactccacc ttggcctcct gagtagctga gactactggc tagcaccacc 7560
 acaccagct ttttattatt attttagtag acaaggtctc aatatgttgc ccaggctagt 7620
 ctcaaaccct tggggtcaag agatcctccg ccctcggcct cccaaagtgc tgggattcca 7680
 ggcattggggc tccgagcccg gcctgcccaa cttaataata cttgttctc agagttgcaa 7740
 ctccaaatga cctgagattg gtgcctttat tctaagctat tttcattttt tttctgtctg 7800
 cattattctc ccccttctct cctccagtct tatctgatat ctgcctcctt cccaccacc 7860
 ctgcacccca tcccaccct ctgtctctcc ctgttctcct caggagactc tggcttctcg 7920
 ttttctcca ctctatctt ttatctctcc ctctacgggt ttcttttctt tctccccggc 7980
 ctgcttgttt ccccccaac ccccttcate tggatttctt cttctgccat tcagtttggg 8040
 ttgagctctc tgcttctccg gtccctctg agctagctgt cccttcaccc actgtgaact 8100
 gggtttccct gcccaaccct cattctctt ctttcttct tttttttttt tttttttttt 8160
 tttttttttt gagacagagt cttgtctctg tgcccagcct ggagtgcagt ggtgcaatct 8220
 tggttcactg caacctccac tcccagatt caagcaatc tctgcctca gcctccagag 8280
 tagctgggat tacaggcgtg tcccaccaca cccgactaat ttttgtattt ttggtagaga 8340
 caaggcttcg gcattgttgg ccaggcaggt ctggaactcc tgacctcaag taatctgcct 8400
 gcctcaccct cccaaagtgc tgggattaca ggcattgagc acctcaccg gaccatccct 8460
 cattctccat cctttcctcc agttgtgatg tctaccctc atgtttccca acaagcctac 8520
 tgggtgtctga atccaggctg ggaagagaag ggagcggctc ttctgtcgga gtctgcacca 8580
 ggccatgct gagacgagag ctggcgctca gagaggggaa gcttggatgg aagcccagga 8640
 gccgcggca ctctcttctc ctcccacccc ctcagttctc agagacgggg aggagggttc 8700
 ccaccaacgg gggacaggct gagacttgag cttgtatctc ctgggccagc tgcaacatct 8760
 gcttgtccct ctgcccactt tggctcctgc acacctgaa cttggtgctt tccctggcac 8820
 tgctctgac acccagctg aggcagcacc cctccctgg agatgactca ccagggtga 8880
 gtgaggagg gaagggtcag tgtgctcaca ggcagggggc ctggtctgct gggcctgctg 8940
 ctgattcacc gtatgtccag gagctggagc agcacctct cagagcctca gtttccccag 9000
 atgtcaatga gagaatcagc atcagccaca tctccacct gaagaatcga accttgagtt 9060
 ctacacttta agaattttt tttttttttg agacagtgcc ttgctcagtc gaccacgtg 9120
 gagtgcagta gcacgatcat caatcactgc agcctccaac tcaagcaacc ctctgcctc 9180
 aacctcctga gtagctgggg ccacaggcac ccaccacat gcctggctaa ttctttcttt 9240
 tttgagatgg agtctcactc tgtcgcccag gctggagtgc agtggcacca tcttggctca 9300
 ccgcaacctc tgccccccgg gttcaagcaa ttctcctgcc tcagcctcct gagtagctag 9360
 aattacaggc gcacaccacc acaccaggct aatttttgta ttttagtag aaacgaggtt 9420
 tcaccacggt ggtcaggttg gtctcaact cctgacatca ggtgatccac ccgcctcggc 9480
 ctcccaaagc actgggatta caggagttag ccaccatgcc tggccatatt ttaaactttt 9540
 ttgtagagat ggggtcttac tatgttgccc aggctgatct caaactcttg ggctcaagca 9600
 gatcctcctg ccaaagcttc ccaaagtagt tggattacag gtgtgagcca ccacgcccc 9660
 ccaacctatg aagatgtcta agcagacagg gtgcctatc tcaccacatc cccagtcaga 9720
 tccagagggt ggacaaggat ggaaggagg ttccagaaca aaggctggca gggagtcgta 9780
 tgggacagga gctgaccag caaccatcca cagagacatc ctggagcctg ggaaggagaa 9840


```

ggacaaagag cccctttttt taaatttttt ttatgttttt gagacggagt ctactctgt 9900
cacccaagct ggagtgcagt ggcacaatct tggctcactg caacctccac ctcatgggtt 9960
caaacgactc tcctgcctca gcctcctgag tagctgggac cacaggtgca caccaccatg 10020
cctggataat ttttgaattt ttggtagaga cggagtttca ccatgttggc caggcaggtc 10080
tcgaactcct gacctcaagt gtcctgcccc ccttggcctc ccaaagttct gggaatacag 10140
gcgtgagcca ctgcaaccag ccagtagccc ccactcttgc ccctcgctga gccctactgg 10200
atgttcttgg ttatgcgaca gtttcccat ctattaaaga gaaacccta tagcagaggg 10260
gaggatgagg ttggaaaagc aggagcattg ttatgctatt cttgtggggt ctgggaagca 10320
gacatctggg tggatgtttg gggggtgctg ggcttagttg gggaagtagg ggggccctg 10380
gggctgacag ggactggaag ctctgagctg gccagaggga tgttgcaatc ctgccagggt 10440
cttgtctatg ctgtcctttt cacaaccatc cccctactgc caggctgaca cgtggttgcg 10500
ggggcacaag gccagccaac ctagagtctg aggctaggcg gaggacacc tccccaccag 10560
ctgccagggt cactggcggt caaaggcagc tgggtgggaa ggcattggac tccagccttg 10620
ggggacggat gtagtgatgg tgggaagcag gcttggtgcc aggaggggag tcagagggtg 10680
aataaaagca gatagagtgt ttgggggagg tagcca 10716

```

<210> 2

<211> 954

<212> DNA

<213> Homo sapiens

<400> 2

```

atgaaggttc tgtgggctgc gttgctggtc acattccttg caggatgcca ggccaagggtg 60
gagcaagcgg tggagacaga gccggagccc gagctgcgcc agcagaccga gtggcagagc 120
ggccagcgct gggaaactggc actgggtcgc ttttgggatt acctgcgctg ggtgcagaca 180
ctgtctgagc aggtgcagga ggagctgctc agctcccagg tcacccagga actgagggcg 240
ctgatggacg agaccatgaa ggagttgaag gcctacaaat cggaactgga ggaacaactg 300
accccggtgg cggaggagac gcgggcacgg ctgtccaagg agctgcaggc ggcgcaggcc 360
cggctgggag cggacatgga ggacgtgtgc ggccgcctgg tgcagtaccg cggcgagggtg 420
caggccatgc tcggccagag caccgaggag ctgcgggtgc gcctcgctc ccacctgcgc 480
aagctgcgta agcggctcct ccgcgatgcc gatgacctgc agaagcgctt ggcaagtgtac 540
caggccgggg cccgcgaggg cgccgagcgc ggccctcagc ccattccgca gcgcctgggg 600
cccctggtgg aacagggccg cgtgcggggc gccactgtgg gctccctggc cggccagccg 660
ctacaggagc gggcccaggc ctggggcgag cggctgcgcg cgcggatgga ggagatgggc 720
agccggaccc gcgaccgcct ggacgaggtg aaggagcagg tggcggaggt gcgcgccaag 780
ctggaggagc agggcccagc gatacgctg caggccgagg ccttccaggc ccgcctcaag 840
agctggttcg agcccctggt ggaagacatg cagcgccagt gggccgggct ggtggagaag 900
gtgcaggctg ccgtgggcac cagcgccgcc cctgtgcccc gcgacaatca ctga 954

```

<210> 3

<211> 317

<212> PRT

<213> Homo sapiens

<400> 3

Met Lys Val Leu Trp Ala Ala Leu Leu Val Thr Phe Leu Ala Gly Cys

1	5	10	15
Gln Ala Lys Val Glu Gln Ala Val Glu Thr Glu Pro Glu Pro Glu Leu	20	25	30
Arg Gln Gln Thr Glu Trp Gln Ser Gly Gln Arg Trp Glu Leu Ala Leu	35	40	45
Gly Arg Phe Trp Asp Tyr Leu Arg Trp Val Gln Thr Leu Ser Glu Gln	50	55	60
Val Gln Glu Glu Leu Leu Ser Ser Gln Val Thr Gln Glu Leu Arg Ala	65	70	75
Leu Met Asp Glu Thr Met Lys Glu Leu Lys Ala Tyr Lys Ser Glu Leu	85	90	95
Glu Glu Gln Leu Thr Pro Val Ala Glu Glu Thr Arg Ala Arg Leu Ser	100	105	110
Lys Glu Leu Gln Ala Ala Gln Ala Arg Leu Gly Ala Asp Met Glu Asp	115	120	125
Val Cys Gly Arg Leu Val Gln Tyr Arg Gly Glu Val Gln Ala Met Leu	130	135	140
Gly Gln Ser Thr Glu Glu Leu Arg Val Arg Leu Ala Ser His Leu Arg	145	150	155
Lys Leu Arg Lys Arg Leu Leu Arg Asp Ala Asp Asp Leu Gln Lys Arg	165	170	175
Leu Ala Val Tyr Gln Ala Gly Ala Arg Glu Gly Ala Glu Arg Gly Leu	180	185	190
Ser Ala Ile Arg Glu Arg Leu Gly Pro Leu Val Glu Gln Gly Arg Val	195	200	205
Arg Ala Ala Thr Val Gly Ser Leu Ala Gly Gln Pro Leu Gln Glu Arg	210	215	220
Ala Gln Ala Trp Gly Glu Arg Leu Arg Ala Arg Met Glu Glu Met Gly	225	230	235
Ser Arg Thr Arg Asp Arg Leu Asp Glu Val Lys Glu Gln Val Ala Glu	245	250	255
Val Arg Ala Lys Leu Glu Glu Gln Ala Gln Gln Ile Arg Leu Gln Ala			

260

265

270

Glu Ala Phe Gln Ala Arg Leu Lys Ser Trp Phe Glu Pro Leu Val Glu
275 280 285

Asp Met Gln Arg Gln Trp Ala Gly Leu Val Glu Lys Val Gln Ala Ala
290 295 300

Val Gly Thr Ser Ala Ala Pro Val Pro Ser Asp Asn His
305 310 315

<210> 4

<211> 15

<212> DNA

<213> Homo sapiens

<400> 4

ggcgtgayta ccgcc

15

<210> 5

<211> 15

<212> DNA

<213> Homo sapiens

<400> 5

ggacgtcytt cccca

15

<210> 6

<211> 15

<212> DNA

<213> Homo sapiens

<400> 6

agctcagrgg cctct

15

<210> 7

<211> 15

<212> DNA

<213> Homo sapiens

<400> 7

ctgggaascc ctggc

15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
gcctccargt agtct

15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
cgggggtrag gcaag

15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
gcagagamga cccga

15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
ggttctgygg gctgc

15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
gctggterca ttcct

15

<210> 13
<211> 15
<212> DNA

<213> Homo sapiens

<400> 13

cattcagrca gaccc

15

<210> 14

<211> 15

<212> DNA

<213> Homo sapiens

<400> 14

ggcccggmtg ggcgc

15

<210> 15

<211> 15

<212> DNA

<213> Homo sapiens

<400> 15

agcggctsct ccgcg

15

<210> 16

<211> 15

<212> DNA

<213> Homo sapiens

<400> 16

gggccgcrtg cgggc

15

<210> 17

<211> 15

<212> DNA

<213> Homo sapiens

<400> 17

cactgaaygc cgaag

15

<210> 18

<211> 15

<212> DNA

<213> Homo sapiens

<400> 18

attacaggcg tgayt

15

<210> 19

<211> 15

<212> DNA

<213> Homo sapiens

<400> 19

gctggggcg gtart

15

<210> 20

<211> 15

<212> DNA

<213> Homo sapiens

<400> 20

ggtgaaggac gtcyt

15

<210> 21

<211> 15

<212> DNA

<213> Homo sapiens

<400> 21

ggctcctggg gaarg

15

<210> 22

<211> 15

<212> DNA

<213> Homo sapiens

<400> 22

gggatgagct cagrg

15

<210> 23

<211> 15

<212> DNA

<213> Homo sapiens

<400> 23

ctttctagag gccyc

15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
gggaccctgg gaasc

15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
ctggaggcca gggst

15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
cccctggcct ccarg

15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
ctcctgagac tacyt

15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
gaggagcggg ggtra

15

<210> 29
<211> 15
<212> DNA

<213> Homo sapiens

<400> 29

ctgctgcttg cctya

15

<210> 30

<211> 15

<212> DNA

<213> Homo sapiens

<400> 30

tgggcagcag agamg

15

<210> 31

<211> 15

<212> DNA

<213> Homo sapiens

<400> 31

agcgggtcgg gtckt

15

<210> 32

<211> 15

<212> DNA

<213> Homo sapiens

<400> 32

gatgaagggtt ctgyg

15

<210> 33

<211> 15

<212> DNA

<213> Homo sapiens

<400> 33

agcaacgcag cccrc

15

<210> 34

<211> 15

<212> DNA

<213> Homo sapiens

<400> 34

tgcgttgctg gtcrc

15

<210> 35

<211> 15

<212> DNA

<213> Homo sapiens

<400> 35

cctgccagga atgyg

15

<210> 36

<211> 15

<212> DNA

<213> Homo sapiens

<400> 36

tggccccatt cagrc

15

<210> 37

<211> 15

<212> DNA

<213> Homo sapiens

<400> 37

ggcccagggt ctgyc

15

<210> 38

<211> 15

<212> DNA

<213> Homo sapiens

<400> 38

ggcgcaggcc cggmt

15

<210> 39

<211> 15

<212> DNA

<213> Homo sapiens

<400> 39

atgtccgcgc ccakc

15

<210> 40
<211> 15
<212> DNA
<213> Homo sapiens

<400> 40
tgcgtaagcg gctsc

15

<210> 41
<211> 15
<212> DNA
<213> Homo sapiens

<400> 41
cggcacgcgc gagsa

15

<210> 42
<211> 15
<212> DNA
<213> Homo sapiens

<400> 42
ggaacagggc cgcr

15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
gtggcgccc gcayg

15

<210> 44
<211> 15
<212> DNA
<213> Homo sapiens

<400> 44
gacaatcact gaayg

15

<210> 45
<211> 15
<212> DNA

<213> Homo sapiens

<400> 45

tgcaggcttc ggcrt

15

<210> 46

<211> 10

<212> DNA

<213> Homo sapiens

<400> 46

acaggcgtga

10

<210> 47

<211> 10

<212> DNA

<213> Homo sapiens

<400> 47

gggggcggta

10

<210> 48

<211> 10

<212> DNA

<213> Homo sapiens

<400> 48

gaaggacgtc

10

<210> 49

<211> 10

<212> DNA

<213> Homo sapiens

<400> 49

tcctggggaa

10

<210> 50

<211> 10

<212> DNA

<213> Homo sapiens

<400> 50

atgagctcag

10

<210> 51

<211> 10

<212> DNA

<213> Homo sapiens

<400> 51

tctagaggcc

10

<210> 52

<211> 10

<212> DNA

<213> Homo sapiens

<400> 52

accctgggaa

10

<210> 53

<211> 10

<212> DNA

<213> Homo sapiens

<400> 53

gaggccaggg

10

<210> 54

<211> 10

<212> DNA

<213> Homo sapiens

<400> 54

ctggcctcca

10

<210> 55

<211> 10

<212> DNA

<213> Homo sapiens

<400> 55

ctgagactac

10

<210> 56
<211> 10
<212> DNA
<213> Homo sapiens

<400> 56
gagcgggggt

10

<210> 57
<211> 10
<212> DNA
<213> Homo sapiens

<400> 57
ctgcttgctt

10

<210> 58
<211> 10
<212> DNA
<213> Homo sapiens

<400> 58
gcagcagaga

10

<210> 59
<211> 10
<212> DNA
<213> Homo sapiens

<400> 59
gggtcgggtc

10

<210> 60
<211> 10
<212> DNA
<213> Homo sapiens

<400> 60
gaaggttctg

10

<210> 61
<211> 10
<212> DNA

<213> Homo sapiens

<400> 61

aacgcagccc

10

<210> 62

<211> 10

<212> DNA

<213> Homo sapiens

<400> 62

gttgctggtc

10

<210> 63

<211> 10

<212> DNA

<213> Homo sapiens

<400> 63

gccaggaatg

10

<210> 64

<211> 10

<212> DNA

<213> Homo sapiens

<400> 64

ccccattcag

10

<210> 65

<211> 10

<212> DNA

<213> Homo sapiens

<400> 65

ccagggctctg

10

<210> 66

<211> 10

<212> DNA

<213> Homo sapiens

<400> 66

gcaggcccg

10

<210> 67

<211> 10

<212> DNA

<213> Homo sapiens

<400> 67

tccgcgcca

10

<210> 68

<211> 10

<212> DNA

<213> Homo sapiens

<400> 68

gtaagcggct

10

<210> 69

<211> 10

<212> DNA

<213> Homo sapiens

<400> 69

catcgcgag

10

<210> 70

<211> 10

<212> DNA

<213> Homo sapiens

<400> 70

acagggccgc

10

<210> 71

<211> 10

<212> DNA

<213> Homo sapiens

<400> 71

gcggcccgca

10

<210> 72
<211> 10
<212> DNA
<213> Homo sapiens

<400> 72
aatcactgaa

10

<210> 73
<211> 10
<212> DNA
<213> Homo sapiens

<400> 73
aggcttcggc

10

<210> 74
<211> 18
<212> DNA
<213> Homo sapiens

<400> 74
tgtaaaacga cggccagt

18

<210> 75
<211> 19
<212> DNA
<213> Homo sapiens

<400> 75
aggaaacagc tatgaccat

19

<210> 76
<211> 10716
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (2836)
<223> PS1: polymorphic base C or T

<220>
<221> allele

<222> (3293)
<223> PS2: polymorphic base C or T

<220>
<221> allele
<222> (3345)
<223> PS3: polymorphic base G or A

<220>
<221> allele
<222> (3376)
<223> PS4: polymorphic base C or G

<220>
<221> allele
<222> (3389)
<223> PS5: polymorphic base G or A

<220>
<221> allele
<222> (3444)
<223> PS6: polymorphic base G or A

<220>
<221> allele
<222> (3492)
<223> PS7: polymorphic base C or A

<220>
<221> allele
<222> (4103)
<223> PS8: polymorphic base T or C

<220>
<221> allele
<222> (4121)
<223> PS9: polymorphic base A or G

<220>
<221> allele
<222> (4211)
<223> PS10: polymorphic base G or A

<220>
<221> allele
<222> (6126)
<223> PS11: polymorphic base C or A

<220>
 <221> allele
 <222> (6150)
 <223> PS12: polymorphic base T or C

<220>
 <221> allele
 <222> (6260)
 <223> PS13: polymorphic base C or G

<220>
 <221> allele
 <222> (6288)
 <223> PS14: polymorphic base C or T

<220>
 <221> allele
 <222> (6384)
 <223> PS15: polymorphic base G or A

<220>
 <221> allele
 <222> (6718)
 <223> PS16: polymorphic base C or T

<400> 76
 ttcattcttac gttagtggaa actgaggctt ccagaggttt caaggtcaca actaaatcca 60
 gaacctcatc tcaggcacac tggtcgtagt cccaatgtcc agtcttaagt cttcttggat 120
 atctgtggct cacagatttt ggggtgttga gcctcctgct gagcactgct ggggccacag 180
 cggtgaccag ccctgtcttc acgggactca gtgagaggaa cagattcatc cgcagagtgg 240
 gcaggactag gttgggggaa cccaggggtc tagagggtt ttcagagggc aggggtcact 300
 gagcggagag cagaggagga gtgagccatt tgctccagcg tgaagttgtt ggtgtgatgg 360
 ggtttcaggg tggcaggagc agtgtgttta aaggctctga agctgtcggc atgtggctgg 420
 tatccaagggt ggccagggaac tctgcatgga tatggtggga agctggcacg cctctcacct 480
 cagctcttcc ctgcaggctc tgtggatagc aactggatcg tgggtgccac gctggagaag 540
 aagctccac ccctgcccc gacactggcc cttggggcct tcctgaatca ccgcaagaac 600
 aagtttcagt gtggcttttg cctcaccatc ggctgagccc tcctggcccc cgccttcac 660
 gcccttcga ttccacctcc acctccacct cccctgccca cagaggggag acctgagccc 720
 ccctcccttc cctccccct tgggggtcgg gggggacatt ggaaaggagg gaccccgcca 780
 cccagcagc tgaggagggg attctggaac tgaatggcgc ttcgggattc tgagtagcag 840
 gggcagcatg cccagtgggc ctgggggtccc gggagggatt ccggaattga ggggcacgca 900
 ggattctgag caccaggggc agaggcggcc agacaacctc agggaggagt gtcctggcgt 960
 ccccatcctc caaagggcct gggcccgcgc cgagggggca gcgagaggag cttccccatc 1020
 cccggtcagt ccacctgcc ccgtccactt tccatctcc tcggtataaa tcatgtttat 1080
 aagttatgga agaaccggga cattttacag aaaaaaaca aaaaacaaca aaaaatatac 1140
 gtgggaaaaa aaacgatggg aggcctccgt tttctcaagt gtgtctggcc tgttttgagc 1200
 atttcatccg gagtctggcc gccctgacct tccccagcc gcctgcaggg ggcgccagag 1260
 ggccggagca cggaaagcag cggatccttg atgtgcctt aagtcgggt cagaggggag 1320

cagcgtggcc tgggggtcgt atcttcccat ccggaacatc tgccctgctg ggggacacta 1380
cgggccttcc cttgcctgag ggtagggctc caaggctact tgccccagc ttgacctggc 1440
cggagtggct atagaggact ttgtccctgc agactgcagc agcagagatg acactgtctc 1500
tgagtgcaga gatgggggca gggagctggg agaggggtca agctactgga acagcttcag 1560
aacaactagg gtactaggaa ctgctgtgtc agggagaagg ggctcaagga ctgcgaggcc 1620
tgggaggagg ggcctaggcc agccatggga gttgggtcac ctgtgtctga ggacttgggtg 1680
ctgtctggat tttgccaaac tagggctggg gtcagctgat gccaccacg actcccgagc 1740
ctccaggaac tgaaccctg tctgccccca gggctctgggg aaggaggctg ctgagtagaa 1800
ccaacccag gttaccaacc ccacctcagc cacccttgc cagccaaagc aaacaggccc 1860
ggcccgccac tgggggttcc ttctcgaacc aggagttcag cctccctga cccgcagaat 1920
cttctgatec caccgctcc aggagccagg aatgagctcc agtctctccc agttctcact 1980
gtgtggtttt gccattcgtc ttgctgctga accacgggtt tctcctctga aacatctggg 2040
atttataaca gggcttagga aagtgcagc gtctgagcgt tcaactgtggc ctgtccattg 2100
ctagccctaa cataggaccg ctgtgtgcca gggctgtcct ccatgctcaa tacacgttag 2160
cttgtcacca aacatacccg tgccgctgct tcccagctc gatgagcaaa ggaacttgat 2220
gtcagagag gacaagtcac ttgccaagg tcacacagct ggcaactggc agagccagga 2280
ttcacgccct ggcaatttga ctccagaatc ctaaccttaa ccagaagca cggcttcaag 2340
cccctggaaa ccacaatacc tgtggcagcc agggggagggt gctggaatct catttcacat 2400
gtggggaggg ggctcccctg tgctcaaggt cacaaccaa gaggaagctg tgattaaaac 2460
ccaggtecca ttgcaaagc ctgcactttt agcagggtgca tcatactgtt cccaccctc 2520
ccatcccact tctgtccagc cgcttagccc cactttcttt ttttctttt tttgagacag 2580
tctccctctt gctgaggctg gagtgcagtg gcgagatctc ggctcactgt aacctccgcc 2640
tcccgggttc aagcgattct cctgcctcag cctcccaagt agctaggatt acaggcgccc 2700
gccaccacgc ctggctaact tttgtatttt tagtagagat ggggtttcac catgttggcc 2760
aggctggtct caaactcctg accttaagt attcgeccac tgtggcctcc caaagtgtctg 2820
ggattacagg cgtgaytacc gccccagcc cctcccatcc cacttctgtc cagcccccta 2880
gccctaactt ctttctggga tccaggagtc cagatcccca gccccctctc cagattacat 2940
tcatccaggc acaggaaagg acagggtcag gaaaggagga ctctgggcgg cagcctccac 3000
attccccttc cacgcttggc cccagaatg gaggagggtg tctgtattac tgggcgaggt 3060
gtcctccctt cctggggact gtggggggtg gtcaaaagac ctctatgcc cacctccttc 3120
ctccctctgc cctgctgtgc ctggggcagg gggagaacag cccacctcgt gactgggggc 3180
tggcccagcc cgccctatcc ctgggggagg gggcgggaca gggggagccc tataattgga 3240
caagtctggg atccttgagt cctactcagc ccagcggag gtgaaggacg tcyttcccca 3300
ggagccggtg agaagcgag tcgggggcac ggggatgagc tcagrggcct ctagaaagag 3360
ctgggaccct gggaascct ggctccarg tagtctcagg agagctactc ggggtcgggc 3420
ttggggagag gaggagcggg ggtraggcaa gcagcagggg actggacctg ggaagggtg 3480
ggcagcagag amgaccgac ccgctagaag gtgggggtgg gagagcagct ggactgggat 3540
gtaagccata gcaggactcc acgagttgtc actatcattt atcgagcacc tactgggtgt 3600
ccccagtgtc ctgagatctc cataactggg gagccagggg cagcgacacg gtagctagcc 3660
gtcgattgga gaactttaaa atgaggactg aattagctca taaatggaac acggcgctta 3720
actgtgaggt tggagcttag aatgtgaagg gagaatgagg aatgcgagac tgggactgag 3780
atggaaccgg cgggtggggag ggggtggggg gatggaattt gaaccccggg agaggaagat 3840
ggaattttct atggaggccg acctggggat ggggagataa gagaagacca ggaggaggtt 3900
aaatagggaa tgggttgggg gcggcttggg aaatgtgctg ggattaggct gttgcagata 3960
atgcaacaag gcttgggaagg ctaacctggg gtgaggccgg gttggggccg ggctgggggt 4020
gggaggagtc ctactggcg gttgattgac agtttctcct tcccagact ggccaatcac 4080
aggcaggaag atgaaggttc tgygggctgc gttgctggtc rcattcctgg caggtatggg 4140
ggcggggctt gctcggttcc cccgctcct cccctctca tctcacctc aacctcctg 4200

```

ccccattcag rcagaccctg ggccccctct tctgaggett ctgtgctgct tccctggctct 4260
gaacagcgat ttgacgctct ctgggcctcg gtttccccca tccttgagat aggagttaga 4320
agttgttttg ttgttggtgt ttgttggtgt tgttttggtt ttttgagatg aagtctcgct 4380
ctgtcgccca ggctggagtg cagtggcggg atctcggtc actgcaagct ccgcctccca 4440
gggccacgcc attctcctgc ctcagcctcc caagtagctg ggactacagg cacatgccac 4500
cacaccgac taactttttt gtatttttcag tagagacggg gtttcaccat gttggccagg 4560
ctggtctgga actcctgacc tcaggtgatc tgcccgtttc gatctcccaa agtgctggga 4620
ttacaggcgt gagccaccgc acctggctgg gagttagagg tttctaatagc attgcaggca 4680
gatagtgaat accagacacg gggcagctgt gatctttatt ctccatcacc cccacacagc 4740
cctgcctggg gcacacaagg aactcaata catgcttttc cgctgggcgc ggtggctcac 4800
ccctgtaatc ccagcacttt gggaggccaa ggtgggagga tcacttgagc ccaggagttc 4860
aacaccagcc tgggcaacat agtgagaccc tgtctctact aaaaatacaa aaattagcca 4920
ggcatggtgc cacacacctg tgcctctcagc tactcaggag gctgaggcag gaggatcgct 4980
tgagccaga aggtcaaggt tgcagtgaac catgttcagg ccgctgcact ccagcctggg 5040
tgacagagca agaccctgtt tataaataca taatgctttc caagtgatta aaccgactcc 5100
ccccctaccc tgcccaccat ggctccaaag aagcatttgt ggagcacctt ctgtgtgccc 5160
ctaggtacta gatgcctgga cggggtcaga aggacctga cccacctga acttgttcca 5220
cacaggatgc caggccaagg tggagcaagc ggtggagaca gagccggagc ccgagctgcg 5280
ccagcagacc gagtggcaga gcggccagcg ctgggaactg gcactgggtc gcttttggga 5340
ttacctgcgc tgggtgcaga cactgtctga gcaggtgcag gaggagctgc tcagctcca 5400
ggtcacccag gaactgaagt gagtgcctcc atcttgccc ttgacctcc tgggtggcgg 5460
ctatacctcc ccaggctcag gtttcattct gccctgtcg ctaagtcttg gggggcctgg 5520
gtctctgctg gttctagctt cctcttccca tttctgactc ctggctttag ctctctggaa 5580
ttctctctct cagctttgtc tctctctctt ccttctgac tcagtctctc acactcgctc 5640
tggtctgtgc tctgtccttc cctagctctt ttatatagag acagagagat ggggtctcac 5700
tgtgttgccc aggttggtct tgaacttctg ggctcaagcg atcctcccgc ctcggcctcc 5760
caaagtgctg ggattagagg catgagccac cttgcccggc ctctagctc cttcttcgtc 5820
tctgcctctg cctctctgat ctgctctctg catctgtctc tgtctccttc tctcggtcctc 5880
tgcccgttcc cttctctccc tcttgggtct ctctggctca tcccctctc gccgcctcc 5940
tcccagccct tctccccgcc tcccactgtg cgacaccctc ccgcctctc ggccgcaggg 6000
cgctgatgga cgagaccatg aaggagttga aggcctacaa atcggaactg gaggaacaac 6060
tgaccccggt ggcggaggag acgcgggcac ggctgtccaa ggagctgcag gcggcgagg 6120
cccgmtggg cgcggacatg gaggacgtgy gcggccgctt ggtgcagtac cgcggcgagg 6180
tgacggccat gctcggccag agcaccgagg agctgcgggt gcgcctcgcc tcccacctgc 6240
gcaagctgcg taagcggcts ctccgcgatg ccgatgacct gcagaagygc ctggcagtg 6300
accaggccgg ggcgcggag ggcgcggagc gcggcctcag cgccatccgc gagcgctgg 6360
ggccccgtgt ggaacagggc cgctgcggg ccgccactgt gggctcctg gccggccagc 6420
cgctacagga gcgggcccag gcctggggcg agcggctgcg cgcgcggatg gaggagatgg 6480
gcagccggac ccgcgaccgc ctggacgagg tgaaggagca ggtggcggag gtgcgcgcca 6540
agctggagga gcaggcccag cagatacgcc tgcaggccga ggccttccag gccgcctca 6600
agagctgggt cgagccccctg gtggaagaca tgcagcgcca gtgggcccgg ctggtggaga 6660
aggtgcaggc tgccgtgggc accagcgccg cccctgtgcc cagcgacaat cactgaaygc 6720
cgaagcctgc agccatgcga cccacgcca cccgtgcct cctgcctccg cgcagcctgc 6780
agcgggagac cctgtccccg cccagccgt cctcctgggg tggaccctag ttaataaag 6840
attaccaag ttacacgat ctgctggcct cccctgtga tttctctaa gcccagcct 6900
cagtttctct ttctgcccac atactggcca cacaattctc agccccctcc tctccatctg 6960
tgtctgtgtg tatctttctc tctgcccctt ttttttttt tagacggagt ctggctctgt 7020
caccaggct agagtgcagt ggcacgatct tggctcactg caacctctgc ctcttgggtt 7080

```

caagcgattc tgetgcctca gtagctggga ttacaggctc acaccaccac acccggttaa 7140
 tttttgtatt tttagtagag acgagctttc accatgttgg ccaggcaggt ctcaaactcc 7200
 tgaccaagtg atccacccgc cggcctccca aagtgtgag attacaggcc tgagccacca 7260
 tgcccggcct ctgcccctct ttctttttta gggggcaggg aaaggtctca cctgtcacc 7320
 cgccatcaca gctcactgca gcctccacct cctggactca agtgataagt gatcctcccg 7380
 cctcagcctt tccagtagct gagactacag gcgcatacca ctaggattaa tttggggggg 7440
 ggggtggtgtg tgtggagatg ggggtctggct ttgttggcca ggctgatgtg gaattcctgg 7500
 gctcaagcga tactcccacc ttggcctcct gagtagctga gactactggc tagcaccacc 7560
 acaccagct ttttattatt attttagtag acaaggtctc aatatgttgc ccaggctagt 7620
 ctcaaaccct tgggtcaag agatcctccg ccatcgacct cccaaagtgc tgggattcca 7680
 ggcattggggc tccgagcccg gcctgcccac ctttaataata cttgttctct agagttgcaa 7740
 ctccaaatga cctgagattg gtgcctttat tctaagctat tttcattttt tttctgtctg 7800
 cattattctc ccccttctct cctccagtct tatctgatat ctgcctcctt cccaccacc 7860
 ctgcacccca tcccaccct ctgtctctcc ctgttctct caggagactc tggcttctg 7920
 ttttctcca cttctatctt ttatctctcc ctctacgggt ttcttttctt tctccccggc 7980
 ctgcttgttt ctccccaac ccccttcac tggatttctt cttctgcat tcaagtttgt 8040
 ttgagctctc tgccttctcg gtccctctg agctagctgt ccttcaccc actgtgaact 8100
 gggtttccct gcccacccct cattctcttt ctttctttct tttttttttt tttttttttt 8160
 tttttttttt gagacagagt cttgctctgt tgcccagcct ggagtgaggt ggtgcaatct 8220
 tgggttactg caacctccac tcccagatt caagcaattc tctgcctca gcctccagag 8280
 tagctgggat tacaggcgtg tcccaccaca cccgactaat ttttgtattt ttggtagaga 8340
 caaggcttcg gcattgttg ccaggcaggt ctggaactcc tgacctcaag taatctgct 8400
 gcctcacct cccaaagtgc tgggattaca ggcattgagc acctcaccg gaccatccct 8460
 cattctccat cctttctctc agttgtgatg tctacccctc atgtttcca acaagcctac 8520
 tgggtgctga atccaggctg ggaagagaag ggagcggctc ttctgtcga gtctgcacca 8580
 ggccatgct gagacgagag ctggcgctca gagaggggaa gcttggtatg aagcccagga 8640
 gccgcggca ctctcttctc ctcccaccc ctcagttctc agagacgggg aggaggggtc 8700
 ccaccaacgg gggacaggct gagacttgag cttgtatctc ctgggccagc tgcaacatct 8760
 gcttgctcct ctgcccctct tggctcctgc acacctgaa cttggtgctt tccctggcac 8820
 tgctctgatc acccacgtgg aggcagcacc cctccctgg agatgactca ccagggtctga 8880
 gtgaggaggg gaagggtcag tgtgctcaca ggcagggggc ctggtctgct gggcctgctg 8940
 ctgattcacc gtatgtccag gagctggagc agcaccctct cagagcctca gtttccccag 9000
 atgtcaatga gagaatcagc atcagccaca tctccacct gaagaatcga accttgagtt 9060
 ctacacttta agaatttttt tttttttttg agacagtgcc ttgctcagtc gaccacgtg 9120
 gagtgagta gcacgatcat caatcactgc agcctccaac tcaagcaacc ctctgcctc 9180
 aacctcctga gtagctgggg ccacaggcac ccaccacct gctgggctaa ttctttcttt 9240
 tttgagatgg agtctcactc tgcgcccag gctggagtgc agtggcacca tcttggctca 9300
 ccgaacctc tgcccccccg gttcaagcaa ttctcctgcc tcagcctcct gagtagctag 9360
 aattacaggc gcacaccacc acaccaggct aatttttgta ttttagtag aaacgaggtt 9420
 tcaccacggt ggtcaggttg gtctcaaaact cctgacatca ggtgatccac ccgcctcggc 9480
 ctcccaaagc actgggatta caggagttag ccaccatgcc tggccatatt ttaaactttt 9540
 ttgtagagat ggggtcttac tatgttgccc aggtgatct caaactcttg ggctcaagca 9600
 gatcctcctg ccaaagcttc ccaaagtagt tggattacag gtgtgagcca ccacgcccc 9660
 ccaacctatg aagatgtcta agcagacagg gtgcctatc tcaccacatc ccagtcaga 9720
 tccagaggtg ggacaaggat ggggaaggag ttccagaaca aaggctggca gggagtcgta 9780
 tgggacagga gctgaccag caaccatcca cagagacatc ctggagcctg ggaaggagaa 9840
 ggacaaagag cccctttttt taaatttttt ttatgttttt gagacggagt ctactctgt 9900
 cacccaagct ggagtgaggt ggcacaatct tggctcactg caacctccac ctcatgggtt 9960

caaacgactc tcctgcctca gcctcctgag tagctgggac cacaggtgca caccaccatg 10020
cctggataat ttttgaattt ttggtagaga cggagtttca ccatgttggc caggcaggtc 10080
tcgaactcct gacctcaagt gctccgcca ccttggcctc ccaaagttct gggaatacag 10140
gcgtgagcca ctgcaaccag ccagtagccc ccatctttgc cctcgtctga gccctactgg 10200
atgttcttgg ttatgcgaca gtttcccat ctattaaaga gaaaccccta tagcagaggg 10260
gaggatgagg ttggaaaagc aggagcattg ttatgctatt cttgtggggt ctgggaagca 10320
gacatctggg tggatgtttg gggggtgctg ggcttagttg gggaagtagg ggggcccctg 10380
gggctgacag ggactggaag ctctgagctg gccagaggga tgttgcaatc ctgccagggt 10440
cttgtctatg ctgtcctttt cacaaccatc cccctactgc caggctgaca cgtgggttgcg 10500
ggggcacaag gccagccaac ctagagtctg aggctaggcg gaggacaccc tccccaccag 10560
ctgccagggt cactggcggg caaaggcagc tggggggaa ggcattggac tccagccttg 10620
ggggacggat gtagtgatgg tgggaagcag gcttgggtgcc aggaggggcg tcagagggtg 10680
aataaaagca gatagagtgt ttgggggagg tagcca 10716

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79234 A3

(51) International Patent Classification⁷: C12Q 1/68

(74) Agent: **FIELD, Gisela, M.**; Genaisance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(21) International Application Number: PCT/US01/12303

(22) International Filing Date: 16 April 2001 (16.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/197,188 14 April 2000 (14.04.2000) US

(71) Applicant (*for all designated States except US*): **GENAISANCE PHARMACEUTICALS, INC.** [US/US]: Five Science Park, New Haven, CT 06511 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
24 January 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CHOI, Julie, Y.** [US/US]: 38 Elizabeth Street, West Haven, CT 06516 (US). **KLIEM, Stephanie, E.** [DE/US]: 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **KOSHY, Beena** [IN/US]: 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **LEE, Helen, H.** [US/US]: 26 Rivendell Drive, Shelton, CT 06484 (US).



WO 01/79234 A3

(54) Title: HAPLOTYPES OF THE APOE GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human apolipoprotein E (APOE) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the APOE gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the APOE gene in an individual are also disclosed. Polynucleotides containing one or more of the APOE polymorphisms disclosed herein are also described.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/12303

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68
US CL : 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ARTIGA et al. Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. FEBS Letters. January 1998, Vol. 421, No. 2, pages 105-108, especially p. 105-106.	1-3
X, P	NICKERSON et al. Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. Genome Research. October 2000, Vol. 10, No. 10, pages 1532-1545, especially Table 1, Figure 2.	1-3
X --- Y	PAIK et al. Nucleotide sequence and structure of the human apolipoprotein E gene. Proc. Natl. Acad. Sci. USA. May 1985, Vol. 82, pages 3445-3449, especially Figures 2-3.	1-2 ----- 3
Y	SAMBROOK et al. Molecular Cloning: a Laboratory Manual, second edition, Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 1989, page 13.20.	3

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

08 September 2001 (08.09.2001)

Date of mailing of the international search report

25 OCT 2001

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Diana B. Johannsen

Telephone No. 703/308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/12303

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-3 as limited to haplotype 1 of Table 5

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/12303

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Groups 1-51, claim(s) 1-3, in part, drawn to methods for haplotyping APOE comprising determining whether the individual has one of the APOE haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4. It is noted that Groups 1-51 correspond to the haplotypes of Table 5 and the haplotype pairs of Table 4, respectively. For example if Group 1 is elected, claims 1-3 will be examined to the extent that they are limited to methods of haplotyping comprising a step of determining whether an individual has the first haplotype of Table 5 of the APOE gene. Upon election of one of the groups, please specify the Table and number of haplotypes requested.

Groups 52-68, claim(s) 4-6, in part, drawn to a method for genotyping the APOE gene. It is noted that Groups 52-68 correspond to determination of the "nucleotide pair" at one of polymorphic sites PS1-PS16 (Groups 52-67), and each of PS1-PS16, as recited in claim 6 (Group 68), respectively.

Groups 69-188, claim(s) 7-9, in part, drawn to methods for haplotyping the APOE gene. It is noted that Groups 69-188 correspond to determination of the nucleotide at two or more polymorphic sites selected from PS1-16. The claims encompass methods requiring identification of 120 possible combinations of polymorphic sites, and Groups 69-188 each correspond to one of these possible combinations, in the order recited in the claims. If applicants elect any of these groups, please specify the two sites to be examined in the method for haplotype.

Groups 189-309, claim(s) 10-11, in part, drawn to a method for predicting a haplotype pair for the APOE gene by identifying an APOE genotype for an individual at two or more polymorphic sites selected from PS1-PS16 (120 possible combinations), as well as each of PS1-PS16. It is noted that the claims encompass methods requiring identification of 121 possible combinations of polymorphic sites, and that Groups 189-309 each correspond to one of these possible combinations, in the order recited in the claim. For example, if Group 189 is selected, claims 10-11 will be examined to the extent that they apply to the combination of PS1 and PS2. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype.

Groups 310-360, claim(s) 12-13, in part, drawn to a method for identifying an association between a trait and one of the 51 haplotypes and haplotype pairs of the APOE gene. Groups 310-360 each correspond to one of the 51 haplotypes and haplotype pairs encompassed by the claims (i.e., the 21 different haplotypes of Table 5, as well as the 30 different haplotype pairs of Table 4). For example if Group 310 is selected, the claims will be examined to the extent that they apply to the first haplotype of Table 5.

Groups 361-376, claim(s) 14-18, in part, drawn to a composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the APOE gene at a polymorphic site selected from PS1-PS16. It is noted that Groups 361-376 correspond to the PS sites recited in the claim in the order recited. For example, Group 361 corresponds to oligonucleotides "for detecting a polymorphism" at PS1.

Group 377, claims 19-20, drawn to a kit comprising a set of oligonucleotides "designed to genotype" each of PS1-PS16.

Groups 378-399, claims 21-22 and 26, in part, drawn to a polynucleotide comprising a sequence which is a polymorphic variant of a reference sequence for the APOE gene or a fragment thereof, wherein the variant comprises "an isogene defined by a haplotype" selected from Table 5. Groups 378-399 correspond to the 22 haplotypes of Table 5, respectively. For example, if Group 378 is selected, the claims will be examined to the extent they are drawn to polynucleotides comprising "an isogene defined by" the first haplotype of Table 5.

Group 400-421, claim(s) 23-24 and 27-28, in part, drawn to recombinant nonhuman organisms comprising one of the "isogenes defined by" a haplotype of Table 5. Groups 400-421 correspond to the 22 haplotypes of Table 5, respectively. For example, if Group 400 is selected the claims will be examined to the extent that they apply to the first haplotype of Table 5.

Groups 422-435, claim(s) 25, in part, drawn a polynucleotide comprising a sequence which is a polymorphic variant of a fragment of the APOE gene that comprises one of the particular nucleotides recited in the claim at the positions recited in the claim. Groups 422-435 correspond to each of the particular PS site nucleotides recited in the claim in the order recited. For example, if Group 422 is selected, the claims will be examined to the extent they are drawn to a fragment comprising "thymine at PS1".

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/12303

display device, an input device, etc., and function in, e.g., methods of electronic sequence comparison. The genome anthologies of Groups 553-574 comprise particular combinations of polymorphisms not required by the molecules of Groups 361-399 and 422-435, and differ structural structurally and functionally from the organisms of Groups 400-421, the polypeptides of Groups 436-457, the antibodies of Groups 458-479, and the computer systems of groups 502-552 for the same reasons set forth above. Accordingly, the products of Groups 361-399 and 422-435, 400-421, 436-457, 458-479, 502-552, and 553-574 differ from each other in structure, function, and effect, and do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature".

Further, the methods of Groups 1-51, 52-68, 69-188, 189-309, 310-360, and 480-501 have different objectives and require different process steps. The methods of Groups 1-51 require steps of identifying haplotypes and haplotype pairs to achieve the objectives of haplotyping. The methods of Groups 52-68 require steps of identifying a single nucleotide on one gene copy to achieve the objective of genotyping. The methods of Groups 69-188 require steps of identifying a nucleotide at a particular combination of sites to achieve the objective of haplotyping. The methods of Groups 189-309 require steps of identifying nucleotides at two polymorphic sites to achieve the objective of "predicting a haplotype pair". The methods of Groups 310-360 requires steps of comparing frequencies of haplotypes in a population to achieve the objective of "identifying an association between a trait" and a haplotype. The methods of Groups 480-501 require steps of assaying for binding activity to achieve the objective of screening for drugs. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the detection or identification of molecules having the same or common special technical feature, for the reasons discussed above.

Continuation of B. FIELDS SEARCHED Item 3:

USPT, DWPI, Medline, CAPus, Lifesci, Scisearch, Embase, Biosis, GenEmbl, Geneseq 0601, Issued, EST
search terms: ApoE, apoprotein E, apolipoprotein E, polymorphism, mutation, allele####, variant, haplotyp####, genotyp####, promoter, regulator###, alu, inventors' names, SEQ ID NO: 76 (nucleotides 2750-2900)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/12303

Groups 436-457, claim(s) 29, in part, drawn to a polypeptide comprising an amino acid sequence which "is a polymorphic variant of a reference sequence" for APOE and which is encoded by "an isogene defined by one of the haplotypes shown in Table 5". Groups 436-457 correspond to the 22 haplotypes of Table 5, respectively.

Groups 458-479, claim(s) 30, in part, drawn to an antibody which binds to a polypeptide of claim 29. Groups 458-479 correspond to antibodies that bind to polypeptides encoded by "an isogene defined by" one of the 22 haplotypes of Table 5, respectively.

Groups 480-501, claim(s) 31, in part, drawn to a method for screening for drugs "targeting the isolated polypeptide of claim 29". Groups 480-501 correspond to methods of screening polypeptides encoded by "an isogene defined by" one of the 22 haplotypes of Table 5, respectively.

Groups 502-552, claim(s) 32, in part, drawn to a computer system comprising polymorphism data wherein the data comprises the haplotypes shown in Table 5 and the haplotype pairs of Table 4. Groups 502-552 correspond to the 21 haplotypes of Table 5 and the 30 haplotype pairs of Table 4, respectively. For example, if Group 502 is selected, the computer system will be examined to the extent that it applies to the first haplotype of Table 5.

Groups 553-574, claim(s) 33, in part, drawn to genome anthologies comprising APOE isogenes having any one of the haplotypes of Table 5. Groups 553-574 correspond to anthologies comprising one of the haplotypes 1-22 of Table 5 in the order shown in Table 5. For example, if Group 553 is selected, a genome anthology comprising haplotype 1 of Table 5 will be examined. The products claimed in Claims 14-22 and 25-26 include fragments of variant sequences, and the claims do not require, e.g., that particular polymorphic sequences be included therein. Accordingly, the claims are sufficiently broad so as to encompass nucleic acid fragments taught in the art. See, e.g., Paik et al (Proc. Natl. Acad. Sci. USA 82:3445-3449 [5/1985]), teaching molecules comprising fragments of the human APOE gene. As the products of Groups 361-399 and 422-435 do not represent a contribution over the prior art, the claims lack a special technical feature that is the same as or that corresponds to a special technical feature of the other claimed inventions. Thus, there is no special technical feature linking the recited Groups, as would be necessary to fulfill the requirement for unity of invention.

It is also noted that each of the present claims has been presented in improper Markush format, as distinct products and distinct methods are improperly joined in the claims. Each polymorphic site and each molecule containing said polymorphic site is structurally and functionally distinct from and has a different special technical feature than each other polymorphic site and molecules containing said site. The chemical structure of each polymorphism and of each molecule containing the same differs from each other polymorphism and molecule. For example, a polynucleotide comprising PS1 is chemically, structurally, and functionally different from a molecule comprising PS2. As the products and methods encompassed by the claims do not share a special technical feature, the distinct products and methods may not properly be presented in the alternative. Accordingly, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed by the claims, and the claims will be examined only as they read upon the invention of the elected group.

It is also noted that the haplotypes, haplotype pairs, and genotypes encompassed by these claims are also distinct from each other and from the single polymorphisms recited in e.g., claims 4-6. For example, a molecule of haplotype 1, comprising a particular combination of polymorphisms, differs chemically, structurally, and functionally from a molecule of haplotype 2 and from a molecule comprising a single polymorphism (e.g., PS1). The special technical feature of each haplotype or genotype is the combination of polymorphisms contained therein, which feature is lacking from and not shared with each other haplotype or genotype or with, e.g., a molecule comprising any single polymorphism set forth in the claims. Similarly, with respect to the pairs of polymorphism of claims 10-11, each combination of polymorphisms differs from each other combination and from each of the other combinations discussed above (i.e., haplotypes, genotypes, and single polymorphic sites). Thus, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby, and the claims will be examined only as they read upon the invention of the elected group.

Further claims Groups 361-399 and 422-435 (polynucleotides, kits, and various compositions), Groups 400-421 (recombinant organisms), Groups 436-457 (polypeptides), Groups 458-479 (antibodies), Groups 502-552 (computer systems) and Groups 553-574 (genome anthologies) are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids of Groups 361-399 and 422-435 are composed of nucleotides and function in, e.g., methods of nucleic acid hybridization or amplification. These groups are directed to different combinations of nucleic acids that are different from one another and may be employed in different methods. The recombinant organisms of Groups 400-421 are complex organisms that are employed in, e.g., animal research methods. Such organisms cannot be employed as, e.g., probes or primers and they differ in both structure and function from the nucleic acids of Groups 361-399 and 422-435. The polypeptides of Groups 436-457 differ in both structure and function from nucleic acids and transgenic organisms. Polypeptides are composed of amino acids linked by peptide bonds and are arranged in complex combinations of, e.g., alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. Polypeptides function in, e.g., enzymatic assays, protein binding assays, purification methods, etc. The antibodies of Groups 458-479 are composed of amino acids linked by peptide bonds, but further require glycosylation and a unique tertiary structure wherein four subunits are associated via disulfide bonds, forming a Y-shaped dimer. Further, antibodies differ from the other products claimed herein in having an immunological function. The computer systems of Groups 502-552 are composed of, e.g., a CPU, a

Form PCT/ISA/210 (second sheet) (July 1998)



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H	A2	(11) International Publication Number: WO 91/13075 (43) International Publication Date: 5 September 1991 (05.09.91)
(21) International Application Number: PCT/FI91/00046 (22) International Filing Date: 15 February 1991 (15.02.91) (30) Priority data: 482,005 16 February 1990 (16.02.90) US (71) Applicant: ORION-YHTYMÄ OY [FI/FI]; Orionintie 1, SF-02100 Espoo (FI). (72) Inventors: SÖDERLUND, Hans ; Salonkitie 19, SF-01940 Espoo (FI). SYVÄNEN, Ann-Christine ; Mannerheimintie 138 A 11, SF-00270 Helsinki (FI). (74) Agent: ORION CORPORATION; Orion Pharmaceutica, Patent Department, P.O. Box 65, SF-02101 Espoo (FI).		(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent). Published <i>Without international search report and to be republished upon receipt of that report.</i>

(54) Title: METHOD AND REAGENT FOR DETERMINING SPECIFIC NUCLEOTIDE VARIATIONS**(57) Abstract**

Detection of variable nucleotide(s) is based on primer extension and incorporation of detectable nucleoside triphosphates. By selecting the detection step primers from the region immediately adjacent to the variable nucleotide, this variation can be detected after incorporation of as few as one nucleoside triphosphate. Labelled nucleoside triphosphates matching the variable nucleotide are added and the incorporation of a label into the detection step primer is measured. The selection of the detection step primer is important to the method according to this invention and is dependent on the nucleotide sequence of interest. The detection step primers are preferably selected so to span the region immediately toward the 3' end from the variable nucleotide to be detected. The method is useful in identifying specific point mutations and genetic variations.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CZ	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

METHOD AND REAGENT FOR DETERMINING SPECIFIC NUCLEOTIDE
VARIATIONS

TECHNICAL FIELD

5

The present invention relates to a method and reagents for determining specific nucleotide variations in a defined polynucleotide region, and to the use of this method in identifying specific point mutations and genetic variations.

10

BACKGROUND OF THE INVENTION

The genetic information of living organisms is carried in the nucleotide sequence of their genome. In the process of gene expression the nucleotide sequence is translated to amino acid sequences, i.e. proteins. Minor changes in the nucleotide sequence, even a single base substitution, may result in an altered protein product. The altered quality or quantity of given proteins changes the phenotype (i.e. the observable characteristics) of the organism or the cell, which for instance may be observed as a development of a disease.

The knowledge of the exact molecular defects causing inherited diseases, as well as predisposition to genetic disorders and cancer is increasing rapidly. The knowledge of the relevance of somatic mutations in malignancies is, however, limited due to the lack of rapid and reliable assay procedures for screening large numbers of samples.

30

Inherited diseases caused by point mutations include sickle cell anemia and β -thalassemias, which are caused by mutations in the β -globin gene. (Antonarakis, 1989, New England J Med, Vol. 320, pp. 153-163) These mutations

generally involve the replacement, insertion or deletion of one to four nucleotides from the sequence of the normal gene. Sickle cell anemia is caused by homozygosity for one unique base pair substitution in the sixth codon of the β -globin gene. (Antonarakis, supra) A large number of mutations in the β -globin gene that can lead to β -thalassemia have been characterized (Antonarakis, supra).

Other known inherited diseases caused by point mutations include α -thalassemia, phenylketonuria, hemophilia, α_1 -anti-trypsin deficiency (Antonarakis, supra) and cystic fibrosis.

Cystic fibrosis is the most common autosomal recessive genetic disorder. It affects about 1/2000 individuals of Caucasian populations and consequently the carrier frequency is about 5%. The recent cloning and genetic analysis of the cystic fibrosis transmembrane regulator (CFTR) gene (Kerem et al., 1989, Science, Vol 245, pp. 1073-1080) has revealed one major mutation, denoted $\Delta F508$, which is a deletion of three nucleotides leading to loss of the phenylalanine at amino acid residue 508. The prevalence of this mutation is on the average 68% in North American and European patient populations, the range being 40 - 88% in reports containing more than 100 CF chromosomes. Because of the high frequency of cystic fibrosis, efficient methods for the screening of carriers and for prenatal diagnosis are needed in the risk group countries.

An example of a polymorphism, which correlates to predisposition to disease is the three-allelic polymorphism of the apolipoprotein E gene. This polymorphism is due to single base substitutions at two DNA loci on the Apo E gene (Mahley, 1988, Science, Vol. 240, pp. 622-630). It may explain as much as 10 % of the individual variations in the serum cholesterol levels. More than 90 % of patients with type III hyperlipoproteinemia are homozygous for one of the Apo E alleles.

The human major histocompatibility complex is a polymorphic system of linked genes located within a conserved region of the genome. The class II genes within the HLA-D (human leukocyte antigen) region encode a series of highly polymorphic alleles (Thomson, 1988, Annu. Rev. Genet., 22:31-50; Morel et al., 1988, Proc. Natl. Acad. Sci. USA, 85:8111-8115; Scharf et al., 1988, Proc. Natl. Acad. Sci. USA, 85:3504-3508). This polymorphism has been shown to be associated with susceptibility to autoimmune diseases, such as insulin-dependent diabetes and pemphigus vulgaris.

The human ras-gene family, which includes the homologous H-, K- and N-ras genes, is one of the potential targets for mutational changes that play a role in human tumorigenesis. Point mutations in either codon 12, 13 or 61 of the ras genes have been shown to convert these genes into transforming oncogenes (Bos et al., 1985, Nature, Vol. 315, pp. 726-730). Somatic point mutations in the N-ras gene have been detected in association with acute myeloid leukemias (AML) (Farr et al., 1988, Proc. Natl. Acad. Sci. USA, 85:1629-1633) and other hemotological malignancies (Neri et al., 1988, Proc. Natl. Acad. Sci. USA, 85:9268-9272). A method for sensitive detection of the N-ras mutations in small quantities of leukemic cells amongst a vast majority of normal cells would constitute a most valuable tool in the follow-up of therapy of AML and other N-ras associated malignancies.

The detection of the specific base changes in the first an second position of codons 12, 13 and 61 of the N-ras gene requires either hybridization with a large number of different oligonucleotide probes, or direct nucleotide sequence determination of the amplified DNA. One critical

point in both approaches is the proportion of cells containing the mutation. Depending on the method of choice a mutation must be present in 5-20% of the analyzed cell population to be detectable.

5

Point mutations and genetic variations in micro-organisms might lead to altered pathogenicity or resistancy to therapeutics. The human immunodeficiency virus (HIV-1) can develop mutants which are resistant to zidovudine (AZT). The resistant virus isolates contain several point mutations, but three mutations seem to be common to all resistant strains: Asp 67 - Asn (GAC - AAC), Lys 70 - Arg (AAT - GAT) and Thr 215 - Phe (ACC - TTC) or Tyr (ACC - TAC) (Larder and Kemp, 1989, Science 246:1155-1158)

15

It would therefore be significant if changes in nucleotide sequences in the genome of living organisms could be determined accurately and with such efficiency and ease that large numbers of samples could be screened. This would afford opportunities for pre- or postnatal diagnosis of hereditary predispositions or diseases and for detection of somatic mutations in cancer. Such a method could also be used for the selection of cells and strains for industrial biotechnology and for plant and animal breeding. Presently available methods suffer from drawbacks limiting their routine use.

25

Polymorphisms or mutations in DNA sequences are most commonly detected by hybridization to allele-specific oligonucleotide (ASO) probes. The nucleotide sequence of the ASO probes is designed to form either a perfectly matched hybrid or to contain a mismatched base pair at the site of the variable nucleotide residues. The distinction between a matched and a mismatched hybrid is based on i) differences in the thermal stability of the hybrids in the conditions used during hybridization or washing (European Patent Publication

35

EP-237362), ii) differences in the stability of the hybrids analyzed by denaturing gradient electrophoresis or iii) chemical cleavage at the site of the mismatch (European Patent Publication EP-329311).

5

Oligonucleotides with 3'ends complementary to the site of the variable nucleotides have been used as allele-specific primers (European Patent Publication EP-332435). The identification of the variable nucleotide is based on the fact that a mismatch at the 3'end inhibits the polymerization reaction. A similar approach is used in oligomer ligation assays, in which two adjacent oligonucleotides are ligated only if there is a perfect match at the termini of the oligonucleotides (European Patent Publication EP-336731).

15

Cleavage of the DNA sequence with restriction enzymes can be utilized for identification of the variation, provided that the variable nucleotide alters, e.g. creates or destroys, a specific restriction site. Nucleotide sequencing is the most informative method for the determination of variable nucleotides.

The methods referred to above are relatively complex procedures, suffering from drawbacks making them difficult to use in routine diagnostics. The use of allele specific oligonucleotide probes requires careful optimization of the reaction conditions separately for each application. Fractionation by gel electrophoresis is required in several of the methods above. Such methods are not easily automatized.

30

SUMMARY OF THE INVENTION

We have now developed an improved method, which allows the detection of nucleotide variations. This method provides several advantages over prior art methods. The method according to this invention comprises few and easily performed procedures. It is especially suited for routine determinations of point mutations and nucleotide variations, such as single mis-matches, deletions, insertions and inversions. The method according to the present invention allows quantification of the proportion of mutated cells in a sample as well as identification of mutations present in as little as 0.5% of the analyzed cell population. Furthermore the complete protocol of the method disclosed is easily automated, which is becoming increasingly important in routine diagnostics.

The method of detection of the variable nucleotide(s) is based on primer extension and incorporation of detectable nucleoside triphosphates in the detection step. By selecting the detection step primers from the region immediately adjacent to the variable nucleotide, this variation can be detected after incorporation of as few as one nucleoside triphosphate.

Labelled nucleoside triphosphates matching the variable nucleotide are added and the incorporation of a label into the detection step primer is measured.

The selection of the detection step primers is important to the method according to this invention and is dependent on the nucleotide sequence of interest. The detection step primers are preferably selected so as to span the region immediately toward the 3' end from the variable nucleotide to be detected. The detection step primers can also be

complementary to a sequence beginning several nucleotides removed from the variable nucleotide. The only limitation concerning the position of the detection step primers is that the sequence between the 3' end of the detection step primer and the variable nucleotide to be detected must not contain a nucleotide residue of the same type as the one to be detected. The detection step primers can equally well be chosen to be complementary to either the coding or non-coding strands of the nucleotide sequence of interest.

The target nucleic acid is preferably enriched by amplification in vitro to allow detection of one single variable nucleotide in the human genome. The present invention advantageously employs an earlier disclosed method for amplification and affinity modification of the target nucleic acid to be detected. In the present method the target nucleic acid containing the variable nucleotide(s) is immobilized on a solid support permitting the removal of the amplification mixture from the target nucleic acid.

The present invention also encompasses individual primer reagents and reagent combinations or kits for practising the method of the invention. While the precise reagents and packaging will depend on the type of nucleotide variations or point mutations to be detected, such a kit will in general include at least one modified amplification primer having an attachment moiety included and at least one detections step primer, but may also include at least one support adapted to immobilize the modified copies of the target nucleic acid and at least one labelled nucleoside triphosphate.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a set-up for a test according to the disclosed method in the case where one variable
5 nucleotide residue (X_1 or X_2) is detected.

Figure 2 illustrates a set-up for a test according to the disclosed method in which case the detection step primer is chosen to span the nucleotide sequence n nucleotides from
10 the variable nucleotide to be detected.

Figure 3 illustrates a set-up for a test according to the disclosed method, in a case where a plurality of mismatches adjacent to each other, is to be detected. In this
15 case the test is performed in two steps, whereby the detection step primer 1 is eluted before step two is performed.

Symbols used in the figures:

20

- X and X' denote nucleotide residues
- The variable nucleotide to be detected is denoted either X_1 , X_2 , X_3 or X_4
- Y_1 , Y_2 , Y_3 and Y_4 denote nucleotides complementary to
25 X_1 , X_2 , X_3 or X_4 , respectively
- dY denotes a deoxyribonucleoside triphosphate
- dY^* and dY^+ denote labelled deoxyribonucleoside triphosphates
- ddY denotes a dideoxyribonucleoside triphosphate
- 30 - ddY^* and ddY^+ denote labelled dideoxyribonucleoside triphosphates
- $*$ and $+$ denote different detectable labels

DETAILED DESCRIPTION OF THE INVENTION

The present invention resides in a method to determine a specific nucleotide variation in a previously defined region of a nucleic acid present in a complex nucleic acid mixture. This defined region containing the nucleotide variation is herein referred to as the "target nucleic acid".

The method according to the invention is based on a primer extension reaction, using at least one detection step primer complementary to the nucleotide sequence 3' from the variable nucleotide to be detected. The detection of the variable nucleotide is performed by detecting a labelled nucleoside triphosphate incorporated by extension of the detection step primer.

Preferably the target nucleic acid is amplified in vitro prior to the detection step to provide a detectable amount of the target nucleic acid. The actual quantity of nucleic acid will vary depending on the labeling moiety employed in the detection step and the sensitivity of analytical techniques for that moiety. In a preferred method according to the present invention, an affinity moiety is introduced into copies of the region of interest of the target nucleic acid. This is conveniently done by a primer dependent amplification reaction using at least one affinity labelled amplification primer. The copies of the target nucleic acid molecules are immobilized on a solid support with the aid of the introduced affinity moiety.

Different set-ups of the test are described below in the description of the preferred embodiments and further illustrated by the examples. It will be evident to one skilled in the art, that the possibilities of constructing the test are not limited to these given examples and that the set-up of the test is determined by the given mutation or nucleotide variation to be detected.

5 a) Introduction of affinity moieties into copies of the target DNA

10 The source of the target nucleic acid (DNA or RNA) suspected to contain the variable nucleotide residue can be any human, animal or plant cell or a microbe. The target nucleic acid can be isolated from biological samples by conventional nucleic acid purification methods, but according to the present invention it is possible to use crude biological samples.

15 The target nucleic acid comprising the variable nucleotide site of interest is preferably multiplied in vitro using one or more primers close to the nucleotide sequence of interest. The term "amplification" as used herein refers to any primer dependent elongation of a nucleotide sequence with
20 the aid of a polymerizing agent. Suitable primer dependent elongation reactions are for example the polymerase chain reaction (Kleppe et al., 1971, J. Mol. Biol. 56:341-361; and U.S. Patent 4,683,202), processes utilizing both primer dependent replication and independent transcription (European
25 Patent Publication EP-329822) or ligation amplification reactions (PCT Patent Publication WO-89/09835).

We have previously developed a convenient method of nucleotide sequencing, which is well adapted for use as a
30 routine diagnostic tool (U.S. Patent Application S.N. 277,643, which is hereby incorporated by reference). This method makes use of the affinity based hybrid collection method disclosed in U.S. Patent Application S.N. 024,604, incorporated herein by reference. Preferably the target DNA
35 to be tested is amplified in vitro prior to sequencing. In this method one of the primers used in the amplification

reaction comprises an attachment moiety such as biotin. The enriched fragment is captured on a solid matrix, to which streptavidin or avidin has been attached. The excess reagents are completely removed by washing the matrix. The captured
5 fragment is rendered single stranded and the chain termination reactions are carried out directly on the matrix. The products of the chain termination reaction are released and the nucleotide sequence is determined after polyacryamide gel electrophoresis. This diagnostically suitable solid
10 phase sequencing method provides several advantages over the prior art methods for determining nucleotide variations described above, but still contains the laborious, expensive and skill-requiring sequencing gel electrophoresis step.

15 According to the present invention the amplification of the target nucleic acid comprising the suspected nucleotide variation is conveniently done by using a modified primer introducing affinity moieties into the copies of the target nucleic acid sequence. In such an amplification one or both
20 of the amplification primers are modified to include attachment moieties. By selecting which amplification primer is modified, it is possible to determine which strand of a double stranded DNA the variable site is detected on. The target sequence of interest is preferably amplified with a
25 suitable number of cycles of this modified polymerase chain reaction.

As a result of this process, copies of the original target nucleic acid sequence, now modified with attachment
30 moieties, are obtained by incorporation of the modified primers into the synthesized polynucleotide molecules.

The term "attachment moiety" as used herein refers to any component having affinity for another component which
35 forms an affinity pair with that other component. For example, biotin - avidin/streptavidin, antigens or haptens -

antibodies, heavy metal derivatives - thiogroups, various polynucleotides such as homopolynucleotides as poly dG - poly dC, poly dA- poly dT and poly dA- poly U, are such affinity pairs. Any component pairs with strong affinity for each other can be used as the affinity pair. Suitable affinity pairs are also found among ligands and conjugates used in immunological methods. An "attachment moiety" is an affinity moiety or a moiety providing a site for the attachment of an affinity moiety.

10

The term "modified amplification primer" as used herein refers to an oligonucleotide primer modified as to contain an attachment moiety. The oligonucleotide primers may be synthesized by standard chemical methods or prepared by recombinant DNA techniques. The essential feature of the amplification primer is that it is able to specifically bind by base pairing to the original target nucleotide sequence of interest at a point in the sequence such that the site of interest will be amplified. Thus the amplification must be complementary to a region of the target nucleotide sequence between the 3' end of the target nucleic acid and the site of interest. The 3' end of the amplification primer is preferably complementary to the target nucleic acid sequence at a point less than 100 residues removed from the site of interest due to limitations in polymerase enzyme processivity. The size of the primers is preferably between 14 and 40 bases, but primers as short as 8 bases or considerably longer than 40 based may be used. The primers are modified with the attachment moieties using chemical or enzymatic or any other method. Preferably the attachment moiety is attached to the 5' end of the primer. Other sites of attachment are also possible, provided that the base pairing property of the primer and its ability to function in an elongation reaction are retained.

30

b) Separation of the target nucleic acid copies.

After amplification, the target nucleic acid copies are separated from the amplification mixture. In the preferred
5 method of the invention, copies of the target nucleic acid molecules containing the attachment moieties are immobilized on a solid matrix with the aid of a complementary attachment site, e.g. the other component of the affinity pair. The matrix is then washed to remove all unbound material, such as
10 nucleotide triphosphates and primers, that could interfere with the reactions during the detection step. The immobilized target nucleic acid is rendered single-stranded either before or after the immobilization step. The immobilization of the modified target nucleic acid copies makes it possible to
15 reuse the target sequence if multiple determinations are to be performed on the same target sequence of interest.

The term "solid matrix" as used herein refers to any format, such as beads, microparticles, the surface of a
20 microtitration well or a test tube, a dipstick or a filter. The material of the matrix may be polystyrene, cellulose, latex, nitrocellulose, nylon, polyacrylamide, dextran or agarose. The only prerequisite for the material is that the attachment site can be attached to it.

25

c) The detection step primer

After separation from the amplification mixture, the
30 single-stranded DNA fragment is allowed to hybridize to a primer, the detection step primer, which is complementary to the nucleotide sequence 3' of the variable nucleotide. This can be carried out in an immobilized state or in solution.

35 The term "detection step primer" as used herein refers

to an oligo- or polynucleotide primer, which functions as the point of initiation for the primer dependent elongation. The detection step primer is selected as to be hybridizable to a nucleotide sequence immediately or closely adjacent to the variable nucleotide to be detected. The detection step primer can be chosen as to be complementary to either the coding or the non-coding strand of a double stranded target, depending on which strand was immobilized in step b) described above. The detection step primer can be modified, e.g. with an affinity moiety as described in section a) above but using a different affinity moiety. Preferably the detection step primer is not modified.

The selection of the detection step primers is determined by the nature of the nucleotide variation to be detected.

In a preferred embodiment of the method according to the present invention, the detection step primer is selected as to be immediately adjacent to the variable nucleotide to be detected.

In another embodiment of the method according to the present invention the detection step primer is selected to be hybridizable to a nucleotide sequence n nucleotide residues away from the variable nucleotide to be detected. The only limitation as to the number n of nucleotide residues between the 3' end of the detection step primer and the variable nucleotide is that there must be no nucleotide residues identical to the one(s) to be detected among these n nucleotide residues.

In another embodiment of the method according to the present invention two or more variable nucleotide residue are identified. In this case it is necessary to design at least two different detection step primers.

The set-up of the test depending on the nucleotide variations to be detected are further described below.

5 d) The detection step primer extension

10 The detection step primer is annealed to the copies of the target nucleic acid and a solution containing one or more nucleoside triphosphates including at least one labelled or modified nucleoside triphosphate, is added together with a polymerizing agent in conditions favoring primer extension. Either labelled deoxyribonucleoside triphosphates (dNTPs) or chain terminating dideoxyribonucleoside triphosphates (ddNTPs) can be used. The polymerizing agent will extend the primer with the nucleoside triphosphate complementary to the variable nucleotide adjacent to the primer. The extended nucleic acid may be immobilized during this detection step primer extension, for example, via affinity-binding of a modified amplified copy, or it may be immobilized afterwards, for example, via an affinity-modified detector primer. In either case, after washing the affinity-matrix, the incorporated label is measured directly on the matrix or after elution.

25 The term "labelled nucleoside triphosphate" as used herein refers to any nucleoside triphosphate, deoxy- or dideoxynucleosidetriphosphate, labelled with a detectable label or modified as to comprise an attachment moiety capable of binding a detectable label. The method according to the invention is not dependent on the label used, although sensitivity will vary depending on the detectability of the label. The only limitation as to the choice of the detectable label is that it will not disturb the incorporation of the labelled nucleoside triphosphate during the polymerization reaction.

30

35

The term "polymerizing agent" as used herein refers to any enzyme which is capable of primer dependent elongation of nucleic acids. Suitable enzymes include, for example, T7 DNA polymerase, T4 DNA polymerase, the Klenow fragment of Escherichia coli DNA polymerase and other suitable DNA polymerases, reverse transcriptase and polymerases from thermophilic microbes such as Thermus aquaticus and Thermus thermophilus.

10

e) Preferred modes of detecting nucleotide variations

Figure 1 illustrates schematically an embodiment of the invention. The target nucleic acid is represented by X's, with X_1 and X_2 representing two alternative nucleotide residues at the site to be investigated. The detection step primer illustrated in Fig. 1 is represented by Y's, and is complementary to the portion of the target sequence starting immediately 3'-ward of the site to be investigated. In practising the invention the detection step primer is hybridized to the target sequence, and a selected nucleoside triphosphate or a mixture of nucleoside triphosphates is added and a chain extension reaction is allowed to proceed under conditions favorable for elongation of the primer.

25

In the simplest embodiment of the invention, the nucleoside triphosphate mixture contains just the labelled dideoxynucleoside triphosphate (ddNTP) corresponding to either X_1 or X_2 (option a in Fig 1). The incorporation of a ddNTP terminates the primer extension reaction after the incorporation of only this one nucleotide, whereafter it is possible to determine the variable nucleotide by detecting the incorporated label of the added nucleotide.

35

This embodiment is especially suitable when the

nucleotide variation to be detected consists of one single point mutation ($X_1 \rightarrow X_2$). The sample can be divided into two parts, whereafter X_1 is detected from one of the samples and X_2 from the other. This allows the identification of heterozygous samples. It is equally possible to add two or more different and differently labelled ddNTPs to an undivided sample (option b in Fig 1). In this embodiment it is possible to determine more than one point mutation occurring at the same site out of one undivided sample ($X_1 \rightarrow X_2, X_3$ or X_4).

It is also possible to use a labelled deoxynucleoside triphosphate (dNTP), which corresponds to the variable nucleotide to be detected and to determine the incorporation of this label. When a labelled dNTP is used, it is advantageous, but not necessary, to add unlabelled ddNTPs corresponding to the other three nucleotide residues (option c in Fig 1). The addition of chain terminating ddNTPs provides a means for preventing incorporation of possibly remaining NTPs from the modification step (a).

Yet another possibility is to use two or more different, differently labelled dNTPs making it possible to detect heterozygotes in an undivided sample. When using more than one labelled dNTP the results may be difficult to interpret if the nucleotide residue x following the variable nucleotide residue in the target sequence is identical to either of the added ones (option d in Fig 1).

Figure 2 illustrates schematically another embodiment of the invention. The detection step primer in Fig. 2 is complementary to a portion of the target sequence starting n nucleotide residues away from the variable nucleotide to be detected. The only limitation as to the number n of nucleotide residues between the 3' end of the detection step primer and the variable nucleotide is that there must be no

nucleotide residues identical to the one(s) to be detected among these n nucleotide residues. In this case the added nucleoside triphosphates comprise unlabelled dNTPs complementary to the n nucleotides between the primer and the variable nucleotide and at least one labelled nucleoside triphosphate depending on the nucleotide variation to be detected. Two or more differently labelled ddNTPs can of course be used as earlier described in the embodiments illustrated by Figure 1 b and d.

10

Figure 3 illustrates schematically yet another embodiment of the method according to the present invention where two or more variable nucleotide residues are identified. In this case it is necessary to design at least two different detection step primers. The first primer is selected as to be complementary to the nucleotide sequence starting immediately adjacent to the first variable nucleotide residue (primer 1 in Fig. 3). Furthermore one or two detection primers are employed spanning also the first variable nucleotide residue detected by primer 1. The test can be performed by dividing the sample into two, whereby primer 1 is added to sample 1 to identify the first variable nucleotide residue. The two other primers (primers 2 and 3 in Fig 3) are added to sample 2 to detect the second variable nucleotide. Due to the immobilization of the target nucleic acid sequence, the identification of two or more subsequent variable nucleotides can be identified from one single undivided sample, by performing the aforementioned detection steps sequentially and including an elution step after the detection of the first variable nucleotide to remove primer 1.

Reagents for use in practising the method of the invention may be provided individually or may be packaged in kit form. For example, kits might be prepared comprising one or more detection step primers and one or more labelled

nucleoside triphosphates, preferably comprising also packaged combinations of one or more affinity labelled amplification primers and corresponding solid support(s).

- 5 The arrangement of the reagents within containers of the kit will depend on the specific reagents involved. Each reagent can be packaged in an individual container, but various combinations may also be possible.

- 10 The present invention is illustrated with the following examples, which are not intended to limit the scope of the invention.

Example 1.

Identification of the apolipoprotein E polymorphism in the codons for amino acid residues 112 and 158 using [³⁵S] as label

Synthesis of Oligonucleotides

Four PCR primers (P1 -P4) and two detection step oligo primers (D1 and D2) were synthesized on an Applied Biosystems 381A DNA synthesizer. The nucleotide sequence of the oligonucleotides, and their location (given as nucleotide numbers) on the apolipoprotein E gene (Apo E) were:

15 P1: 5'-AAG GAG TTG AAG GCC TAC AAA T (3616 - 3637) [SEQ.ID.NO.1]
P3: 5'-GAA CAA CTG AGC CCG GTG GCG G (3649 - 3670) [SEQ.ID.NO.2]
D1: 5'-GCG CGG ACA TGG AGG ACG TG (3725 - 3744) [SEQ.ID.NO.3]
D2: 5'-ATG CCG ATG ACC TGC AGA AG (3863 - 3882) [SEQ.ID.NO.4]
P2: 5'-TCG CGG GCC CCG GCC TGG TAC A (3914 - 3893) [SEQ.ID.NO.5]
20 P4: 5'-GGA TGG CGC TGA GGC CGC GCT C (3943 - 3922) [SEQ.ID.NO.6]

A 5'-aminogroup was added to the primer P2 with the aminolink II reagent (Applied Biosystems). The amino group was biotinylated using sulfo-NHS-biotin (Pierce Chemical Co.)
25 and purified by reversed phase HPLC.

The DNA Samples

30 Venous blood samples were obtained from patients of known Apo E phenotype attending the Lipid Outpatient Clinic of the University Central Hospital of Helsinki, Finland. Leukocytic DNA was extracted according to standard procedures.

Polymerase chain reaction-Amplification

The DNA (100 ng per sample) was amplified with the P1 and P4 primers (final concentration 1 μ M) in 100 μ l of a solution of 0.2 mM each of dATP, dCTP, dGTP, dTTP, 20 mM Tris-HCl, pH 8.8, 15 mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgCl_2 , 0.1 % Tween 20, 0.1 mg/ml gelatin and 2.5 units of Thermus aquaticus DNA-polymerase (United States Biochemical Corp.) in a DNA thermal cycler (Perkin-Elmer /Cetus) for 25 cycles of 1 min. at 96°C and 2 min. at 65°C. An aliquot (3 μ l of a 1:100 dilution) of this first PCR amplification mixture was transferred to a second PCR. This was carried out at the conditions described above and directed by the biotinylated primer P2 and the primer P3.

15

Affinity-capture of the Biotinylated Amplified Apo E DNA on Avidin-coated Polystyrene Particles

Five μ l of a 5 % (w/v) suspension of avidin-coated polystyrene particles (0.8 μ m, Baxter Healthcare Corp.) were added to an 80 μ l aliquot of the second amplification mixture. The samples were kept at 20°C for 30 min. The particles were collected by centrifugation for 2 min. in an Eppendorf centrifuge and were washed once by vortexing with 1 ml of 15 mM NaCl, 1.5 mM Na-citrate (0.1 x SSC), 0.2 % sodium dodecyl sulphate (SDS), and once with 1 ml of 0.1 % Tween 20 in 0.15 M NaCl, 20 mM phosphate buffer, pH 7.5 (PBS). The particles were treated twice with 200 μ l of 0.15 M NaOH for 5 min. at 20°C. The particles were then washed once with 1 ml of 0.1 % Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5 and twice with 1 ml of 0.01 % Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5. The suspension of particles in the last washing solution was divided into four parts, and the particles were collected by centrifugation in separate tubes.

Identification of the Variable Nucleotides

The particles carrying the DNA fragment were suspended in 10 μ l of 50 mM NaCl, 20 mM MgCl₂, 40 mM Tris-HCl, pH 7.5, containing 2 pmol of the detection step primer. The D1 oligonucleotide located immediately adjacent to the variable nucleotide number 3745 (codon 112, T or C,) was added to two of the tubes and the D2 oligonucleotide adjacent to the variable nucleotide number 3883 (codon 158, T or C) to two tubes. The oligonucleotide was annealed to the DNA template by heating the samples at 65°C for 2 min. and allowing them to cool to 20°C. One μ l of 0.1 M dithiotreitol and [³⁵S]-labelled deoxynucleoside triphosphates (dNTP) and dideoxynucleoside triphosphates (ddNTP) were added to yield 1 μ M concentrations each in a final volume of 15 μ l as follows:

- for identification of T: [³⁵S]-dTTP (Amersham), ddCTP and ddGTP to two tubes, one in which the oligonucleotide D1 and one in which the oligonucleotide D2 had been annealed.
- for identification of C: [³⁵S]-dCTP (Amersham), ddTTP, and ddGTP to two tubes, one in which the oligonucleotide D1 and one in which the oligonucleotide D2 had been annealed.

Two μ l (3U) of T7 DNA polymerase (Sequenase™, United States Biochemical Corp.) was added to each tube and the reaction was allowed to proceed for 6 min. at 42°C. The microparticles were washed at 20°C by vortexing twice with 1 ml of 0.1 x SSC, 0.2% SDS and twice with 0.1 % Tween 20 in PBS. For elution of the reaction products the particles were boiled in 200 μ l of H₂O for 5 min., cooled on ice and centrifuged for 2 min. in an Eppendorf centrifuge. The eluted radioactivity was measured in a liquid scintillation counter. The result of an experiment where four samples of the phenotypes E2/E2 (T/T in codon 112, T/T in codon 158), E4/E3 (T/C in codon 112, C/C in codon 158), E2/E3 (T/T in codon

112, T/C in codon 158) and E4/E4 (C/C in codon 112, C/C in codon 158) were analyzed as described above is presented in the following table:

5

Sample no.	Phenotype	Detector (codon)	Radioactivity eluted (cpm)		Result
			T-reaction	C-reaction	

10

1.	E2/E2	D1 (112)	18400	625	T/T
2.	E4/E3	D1 (112)	73700	58100	T/C
3.	E2/E3	D1 (112)	112000	1360	T/T
4.	E4/E4	D1 (112)	2310	99700	C/C
15 5.	No DNA	D1 (112)	429	1195	

20

1.	E2/E2	D2 (158)	67000	7000	T/T
2.	E4/E3	D2 (158)	3220	35100	C/C
3.	E2/E3	D2 (158)	44600	26400	T/C
20 4.	E4/E4	D2 (158)	1485	19300	C/C
5.	No DNA	D2 (158)	686	760	

Conclusion

The differences in cpm-values obtained in the T- and C-reactions allowed unequivocal identification of the variable nucleotide in both codon 112 and codon 158 in all four DNA samples.

30

The variable nucleotide can optionally be determined as described above, but performing the second PCR with biotinylated primer P3 and primer P2. As detection step primers are then used primers complementary to the opposite

35 strand of the ApoE gene:

D4: 5'- GTA CTG CAC CAG GCG GCC GC (3765 - 3746) [SEQ.ID.NO. 7]

D5: 5'- GGC CTG GTA CAC TGC CAG GC (3903 - 3884) [SEQ.ID.NO. 8]

Example 2

Identification of the variable nucleotide in codon 112 of the apolipoprotein E gene using double labelling ($[^3\text{H}]$ and $[^{32}\text{P}]$) in one reaction.

Oligonucleotides and DNA Samples

The PCR primers P1, biotinylated P2, P3 and P4, and the detection step primer D1 described in Example 1 were used in this example. The DNA was extracted from blood samples as described in Example 1.

Polymerase chain reaction-Amplification and Affinity-capture

The PCR amplification and the affinity-capture of the amplified fragments were carried out as described in Example 1. In the last washing step each sample was divided into two aliquots.

Identification of the variable nucleotide at codon 112

The particles were suspended in 10 μl of buffer (see Example 1) containing 2 pmol of the detection step primer D1, which hybridizes immediately 3' to the variable nucleotide in codon 112. The annealing reaction was carried out as in Example 1. One μl of 0.1 M dithitritol was added.

The identification of the variable nucleotide (C or T) was now done by adding $[^3\text{H}]$ - and $[^{32}\text{P}]$ -labelled dNTPs simultaneously to one sample. Either $[^3\text{H}]$ -dCTP and $[^{32}\text{P}]$ -dTTP or $[^3\text{H}]$ -dTTP and $[^{32}\text{P}]$ -dCTP (Amersham) were used. The $[^3\text{H}]$ -dNTPs were added to 1 μM concentrations. The $[^{32}\text{P}]$ -dNTPs

were diluted in unlabelled dNTP to yield 1 μ M final concentrations and specific activities similar to those of the [3 H]-dNTPs. All reactions contained 1 μ M ddGTP. The final volume was 15 μ l. Three units of T7 DNA polymerase was added, and the labelling and washing procedures were carried out as in Example 1. The eluted [3 H] and [32 P] radioactivity in each sample was measured simulataneously in a Rackbeta 1219 scintillation counter (Pharmacia/Wallac) by setting the window for [3 H] at channels 10-90 and the windows for [32 P] at channels 130-220.

The table below shows the result of an experiment, in which three samples, one of the phenotype E2/E3 (T/T in codon 112), one of the phenotype E4/E4 (C/C in codon 112) and one of the phenotype E3/E4 (T/C on codon 112) were analyzed either using [3 H]-dCTP and [32 P]-dTTP or using [3 H]-dTTP and [32 P]-dCTP:

Sample	Labelled dNTPs	Radioactivity eluted (cpm)	
		3 H; Ch.10-90	32 P; Ch.130-220
E2/E3	[3 H]dCTP / [32 P]dTTP	502	7870
E4/E4	[3 H]dCTP / [32 P]dTTP	6070	186
E3/E4	[3 H]dCTP / [32 P]dTTP	5120	5980
No DNA	[3 H]dCTP / [32 P]dTTP	172	148
E2/E3	[3 H]dTTP / [32 P]dCTP	10800	183
E4/E4	[3 H]dTTP / [32 P]dCTP	394	4932
E3/E4	[3 H]dTTP / [32 P]dCTP	7800	5140
No DNA	[3 H]dTTP / [32 P]dCTP	175	44

35

Conclusion

The signals obtained allowed the identification of the variable nucleotide in codon 112 from undivided samples using two dNTPs carrying different labels. In this example only half of each sample was analyzed per reaction. Thus the other half of the sample can be used to analyze the nucleotide variation in codon 158 of the Apo E gene.

Example 3.

Identification of the variable nucleotide in codon 158 of the apolipoprotein E gene after a single polymerase chain reaction amplification.

Oligonucleotides and DNA samples

In this example the biotinylated primer P2 and the primer P3 were used in the PCR amplification. The detection step primer was D2 (see Example 1). The DNA was extracted from venous blood as described in Example 1.

15

Polymerase chain reaction-Amplification and Affinity-capture

The DNA (100 ng per sample) was amplified with the biotinylated P2 and the P3 primer (final concentration 1 μ M) at the conditions described in Example 1 with the following exception: Only one amplification process consisting of 30 cycles of 1 min. at 96°C, 1 min. at 55°C and 1 min. 72°C was carried out.

25

The affinity-capture on avidin-coated polystyrene particles was done as in Example 1. Each sample was divided into two parts in the last washing step.

30

Identification of the variable nucleotide in codon 158

The particles were suspended in 10 μ l of buffer (see Example 1) containing 2 pmol of the detection step primer D2, which hybridizes immediately 3' of the variable nucleotide in

codon 158. The annealing reaction was as in Example 1. One μ l of 0.1 M dithiotreitol was added. [35 S]-labelled dNTPs and ddNTPs were added to the T- and C-reactions as specified in Example 1. The primer extension reaction, the washing
5 procedure and the elution of the bound radioactivity was done as described in Example 1.

Three samples with the phenotypes E2/E2 (T/T in codon 158), E2/E3 (T/C in codon 158) and E4/E3 (C/C in codon 158),
10 respectively, were analyzed. The result of this experiment is presented in the table below:

Sample	Radioactivity eluted (cpm)		Result
	T-reaction	C-reaction	
E2/E2	64600	7746	T/T
E2/E3	39600	22700	T/C
E4/E3	5640	53500	C/C
No DNA	1325	1910	

Conclusion

25 The method allowed correct identification of the variable nucleotide at position 158 also when the apo E DNA fragment was enriched by a single PCR with one primer pair.

Example 4

Identification of the variable nucleotide in codon 112 of the apolipoprotein E gene with enzymatic detection.

5

Oligonucleotides and DNA samples

The PCR primers P1, biotinylated P2, P3 and P4, and the detection step primer D1 described in Example 1 were used. The DNA was extracted from blood samples as described in Example 1.

15 Polymerase chain reaction-Amplification

The DNA (100 ng per sample) was amplified with the primers P1 and P4 at 1 μ M concentration as described in example 1. Three μ l of a 1:100 dilution of this first PCR mixture was reamplified using the biotinylated P2 and the P3 primer at 0.1 μ M concentration. The second PCR was carried out for 30 cycles of 1 min. at 96°C, 1 min. at 55°C and 1 min. at 72°C.

25

Affinity-capture of the Amplified DNA in Avidin-coated Microtitration wells

Two 15 μ l aliquots per sample of the second PCR mixture were transferred to microtitration wells (Nunc, Maxisorb), that had been coated with streptavidin by passive absorption. 30 μ l of 0.1 % Tween 20 in 0.15 M NaCl, 0.1 M Tris-HCl, pH 7.5 (TBS) was added to each well. The microtitration strips were incubated for 3 hours at 37°C with gentle shaking. The

wells were washed three times with 200 μ l of 0.1 % Tween 20 in TBS at 20°C. The wells were then treated twice with 100 μ l of 50 mM NaOH for 5 min. at 20°C followed by washing twice with 200 μ l of 0.1 x SSC, 0.2 % SDS, twice with 0.1 % Tween 20 in TBS, once with 0.1 % Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5, and finally once with 0.01 % Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5.

10 Identification of the variable nucleotide

Ten pmol of the oligonucleotide D1 was added to each well in 50 μ l of 0.9 M NaCl, 0.2 M Tris-HCl, pH 7.5. The wells were heated at 65°C for 2 min. and allowed to cool slowly to 20°C. The mixture was discarded and the wells were washed once with 200 μ l of 0.25 M NaCl, 0.2 M Tris-HCl, pH 7.5 at 20°C. 50 μ l of a solution consisting of 1 μ M digoxigenin-11-dUTP (Boehringer-Mannheim), 1 μ M ddCTP, 1 μ M ddGTP, 0.2 μ M oligonucleotide D1, 6 mM dithiotreitol, 37.5 mM NaCl, 15 mM MgCl₂, 30 mM Tris-HCl, pH 7.5 and 3 units of T7 DNA polymerase was added. The microtitration strips were incubated at 42°C for 10 min. and the wells were washed twice with 200 μ l of 0.1 x SSC, 0.2 % SDS, and three times with 200 μ l of 0.1 % Tween 20 in TBS. Then 60 μ l of a 1:500 dilution of an anti-digoxigenin-alkaline phosphatase conjugate (Boehringer-Mannheim) in a solution of 0.1 % Tween 20, 1 % bovine serum albumin in TBS was added, and the microtitration strips were incubated at 37°C for 2 hours with gentle shaking. The wells were washed six times with 0.1 % Tween 20 in TBS and once with 1M diethanolamine-0.5 M MgCl₂ buffer, pH 10. Finally 60 μ l of 2 mg/ml p-nitrophenyl phosphate in the alkaline buffer was added. After development of colour for 20 min. at room temperature 100 μ l of the alkaline buffer was added and the absorbance of the formed product was measured at 405 nm in a spectrophotometric reader.

Two samples with the phenotypes E2/E2 (T/T in codon 112) and E4/E4 (C/C in codon 112) were analyzed. The result of this experiment is presented in the table below:

5

Sample	Absorbance at 405 nm (duplicate samples)		Result
E2/E2	1.180	0.707	T/T
E4/E4	0.040	0.010	C/C
No DNA	0.025	0.010	

Conclusion

The variable nucleotide in codon 112 was identified after incorporation of digoxigenin-11-dUTP and subsequent detection with an antibody labelled with alkaline
20 phosphatase.

Example 5.

25 Identification of the apolipoprotein E polymorphism in the codon for amino acid residue 112 using fluorescent labels

Oligonucleotides and DNA samples

30

The PCR primers P1, biotinylated P2, P3 and P4, and the detection step primer D1 described in Example 1 are used. The DNA is extracted from blood samples as described in Example 1.

Polymerase chain reaction-Amplification and Affinity-capture

The DNA (100 ng per sample) is amplified with the primers P1 and P4 at 1 μ M concentration as described in Example 1. Three μ l of a 1:100 dilution of this first PCR mixture is reamplified using the biotinylated P2 and the P3 primer at 1 μ M concentration. The second PCR is carried out for 25 cycles of 1 min. at 96°C, 1 min. at 55°C and 1 min. at 72°C. The biotinylated amplified DNA fragments are captured on avidin-coated polystyrene particles as in Example 1. Each sample is divided into two parts in the last washing step.

Identification of the Variable Nucleotide in Codon 112

The particles carrying the amplified DNA are suspended in 10 μ l of buffer (see Example 1) containing 5 pmol of the detection step primer which hybridizes immediately 3' of the variable nucleotide in codon 112. The annealing reaction is carried out as in Example 1. One μ l of 0.1 M dithiotreitol is added to each tube. For identification of T, 400 pm of fluorescent ddTTP (T-terminator; DuPont, NEK-528T) is added to one of the tubes. For identification of C, 40 pm of fluorescent ddCTP (C-terminator; DuPont, NEK-519C) is added to the other tube. Five units of T7 DNA polymerase is added to both tubes to a final reaction volume of 15 μ l. The reaction is allowed to proceed for 5 min. at 37°C. The particles are washed as described in Example 1 and the reaction products are eluted. The fluorescence of the eluant is measured in a fluorescence spectrophotometer (Merck/Hitachi, F-1000) using 490 nm as the excitation wavelength and 550 nm for measurement of the emitted fluorescence.

Interpretation of the Result

5 A positive signal from only the T-reaction shows that the subject is homozygous for a cystein residue at position 112.

A positive signal from only the C-reaction shows that the subject is homozygous for an arginine residue at position 112.

10 Positive signals from both reactions show that the subject is heterozygous, ie. has one allele with a cystein residue, and one allele with an arginine residue at position 112 of the apolipoprotein E gene.

15 Example 6.

Detection of the sickle cell muatation in the sequence encoding codon 6 of the human β -globin gene.

20

Synthesis of Oligonucleotides

25 The PCR primers are designed to contain 3'-ends that are mismatched to the otherwise strongly homologuous δ -globin gene. Two PCR primers, denoted B1 and B2, and one detection step primer B3, are synthesised by the method described in Example 1. The primer B2 is biotinylated as described in Example 1. The nucleotide sequence of the oligonucleotides and their location on the β -globin gene (as nucleotide
30 numbers relative to the transcription intiation site) are the following:

B1: 5'-CAT TTG CTT CTG ACA CAA CT (-49 - -30) [SEQ.ID.NO. 9]
B3: 5'-CAT GGT GCA CCT GAC TCC TG (-1 - 19) [SEQ.ID.NO. 10]
35 B2: 5'-CAA CTT CAT CCA CGT TCA CC (73 - 54). [SEQ.ID.NO. 11]

Polymerase chain reaction-Amplification and Affinity-capture

The DNA is extracted from blood samples as described in Example 1. The DNA (100 ng per sample) is amplified with the primers B1 and biotinylated B2 as described in Example 3. The biotinylated amplified DNA fragments are captured on avidin-coated polystyrene particles as in Example 1. Each immobilized sample is divided into two parts.

10

Identification of the A → T Mutation in Codon 6

The particles carrying the amplified DNA sample are suspended in 10 μ l of buffer (see Example 1) containing 2 pmol of the B3 detection step primer, which hybridizes immediately 3' of the mutation site in codon 6. The annealing reaction is carried out as in Example 1. One μ l of 0.1 M dithiotreitol is added. [35 S]-labelled dNTPs and ddNTPs are added to yield 0.2 μ M concentrations in a final volume of 15 μ l as follows:

20

- for identification of the normal allele (A):
[35 S]-dATP, ddTTP, ddGTP to one of the tubes
- for identification of the mutation (T): [35 S]-dTTP, ddATP, ddGTP to the other tube.

25

One unit of T7 DNA polymerase is added, and the reaction is allowed to proceed for 5 min. at 37°C. The particles are washed, the reaction products are eluted and the eluted radioactivity is measured in a scintillation counter as described in Example 1.

30

Interpretation of the Result

A positive signal from only the A-reaction shows that the subject is homozygous and normal.

5 A positive signal from only the T-reaction shows that the subject is homozygous for the sickle cell mutation.

A positive signal from both reactions show that the subject carries the sickle cell mutation in one allele ie. is heterozygous.

10

The variable nucleotide can optionally be determined as described above, but performing the PCR with biotinylated primer B1 and primer B2. As detection step primer is then used a primer complementary to the opposite strand of the

15 β -globin gene:

B4: 5'- CAG TAA CGG CAG GCG GCC GC (40 - 21) [SEQ.ID.NO. 12]

20 Example 7.

Identification of the Δ F508 deletion in the cystic fibrosis.

25

Synthesis of oligonucleotide primers

Two amplification primers (CF1 and CF3) and one detection step primer (CF2) were synthesised as described in Example 1. The primers were designed based on the known
30 nucleotide sequence of the CF gene (Riordan et al., Science 1989; 245: 1066-1072). The sequence and the position on the CF gene of the primers were:

CF1: 5'-CTG GAG CCT TCA GAG GGT AAA AT-3' (1555-77) [SEQ.ID.NO.13]
CF2: 5'-TGG CAC CAT TAA AGA AAA TAT CAT-3' (1629-52) [SEQ.ID.NO.14]
CF3: 5'-CAT GCT TTA ATG ACG CTT CTA TA-3' (1703-81) [SEQ.ID.NO.15]

5 The primer CF3 was biotinylated as described in Example
1.

The DNA samples

10

The target DNA was extracted from leukocyte cells from
finnish cystic fibrosis patients, who had been clinically
characterized in regard of the $\Delta F508$ mutation, according to
standard methods.

15

PCR-Amplification and Affinity-capture of the Target DNA

20 The DNA (100 - 200 ng per sample) was amplified with the
biotinylated CF3 and the CF1 primer (final concentrations
0.15 μM and 0.6 μM , respectively) at the conditions decribed
in Example 1, with the following exception: Only one
amplification process consisting of 31 cycles of 1 min at
96°C, 1 min at 60°C and 1 min at 72°C was carried out.

25

The affinity capture on steptavidin-coated
microtitration wells was performed as described in Example 4.

30 Identification of the variable nucleotide

The detection step primer extension was performed as
described in 50 μl of 50 mM Tris-HCl, pH 8.8, 1.5 mM $MgCl_2$,

15 mM $(\text{NH}_4)_2\text{SO}_4$, 0.1% Tween 20, 0.01% gelatin containing 0.2 μM of the primer CF2 and 2 units of Taq DNA polymerase. For identification of T, 1 μCi ^3H -dTTP (Amersham) and 0.8 μM ddGTP, and for identification of C, 2 μCi ^3H -dCTP, 0.8 μM ddTTP were added in two parallel wells and the plates were incubated for 10 min at 55°C. The wells were washed three times with 200 μl of 0.1 % Tween 20 in TBS buffer and the incorporated label was eluted with 60 μl of 50 mM NaOH for 5 min at room temperature and measured in a liquid scintillation counter.

Three samples with the genotypes C/C, T/T and T/C in nucleotide 1653 of the CFTR-gene, i.e. one normal homozygote, one ΔF508 homozygote and one heterozygote, were used in this example. These samples had been previously genotyped by other methods (Kere et al, Hum. Genet., 1990; 85:413-415). The results in the table below show that the identification of each genotype is clear.

Sample	^3H -dTTP (cpm)	^3H -dCTP (cpm)	cpmC/cpmT	Geno- type ^{x)}
Normal homozygote	349	9832	28.2	C/C
ΔF508 homozygote	11539	52	0.005	T/T
Hetero- zygote	11358	7457	0.66	T/C

x) nucleotide 1653 of the CFTR gene.

Conclusion

The genotypes of the samples were verified according to cpmC/cpmT ratios after incorporation of ³H-dTTP and ³HdCTP.

5

Example 8.

10 Detection of point mutations in the sequence encoding codon 12 in the K-ras gene.

Synthesis of oligonucleotides

15

Two PCR primers, denoted R1 and R2, respectively, are synthesised and the primer R1 is biotinylated as described in Example 1. For detection of a mutation of the glycine residue (encoded by GGT) in position 12, two detection step primers, 20 R3 and R4, are synthesised. R3 is used to detect a mutated G in the second position and R4 to detect a mutated G in the first position of codon 12. The sequence and position (as nucleotide numbers) of the oligonucleotides on the first exon of the K-ras gene is given below:

25

R1: 5'-ATG ACT GAA TAT AAA CTT GTG (1 - 20) [SEQ.ID.NO. 16]

R2: 5'-TTC GTC CAC AAA ATG ATT CTG (94 - 74) [SEQ.ID.NO. 17]

R3: 5'-AAG GCA CTC TTG CCT ACG CCA (56 - 36) [SEQ.ID.NO. 18]

R4: 5'-AGG CAC TCT TGC CTA CGC CAC (55 - 35) [SEQ.ID.NO. 19]

30

Polymerase chain reaction-Amplification, Affinity-capture and Annealing of the Detection step Primers

35 The DNA is extracted from tumor cell samples by standard

methods using digestion with proteinase K, phenol extraction and precipitation with ethanol. 100 ng of the purified DNA is amplified with the biotinylated primer R1 and the primer R2 using the conditions described in Example 3, except that the
5 primer annealing temperature is 50°C. The amplified DNA is captured on avidin-coated polystyrene particles, denatured and the particles are washed as described in Example 1. The immobilized DNA sample is divided into two tubes. The particles in one of the tubes are suspended in 10 μ l of
10 buffer (see Example 1) containing 2 pmol of the oligonucleotide R3, which will anneal immediately 3' of the C complementary to the second G in codon 12. In the other tube 10 μ l of buffer containing 2 pmol of the oligonucleotide R4 annealing 3' to the C complementary to the first G in codon
15 12 is used. The annealing reaction is carried out as in Example 1. One μ l of 0.1 M dithiotreitol is added. The mutation in codon 12 is analyzed according to method A or method B below.

20 A. Identification of a gly \rightarrow non-gly Mutation in Codon 12

A mixture of [35 S]-dATP, [35 S]-dGTP, [35 S]dTTP and ddCTP are added to both tubes to give a final concentration of 1 μ M each in 15 μ l. One unit of T7 DNA polymerase is added and the
25 reaction is allowed to proceed for 5 min. at 37°C. The particles are washed, the reaction products are eluted and the eluted radioactivity is measured in a scintillation counter as described in Example 1.

30 A positive signal from the tube, in which R3 had been annealed, shows that at least a part of the K-ras genes in the sample has mutated to a valine (encoded by GTT), an aspartic acid (GAT) or an alanine (GCT) residue in position 12.

A positive signal from the tube, in which R4 had been annealed, shows that there is a cysteine (TGT), a serine (AGT) or an arginine (CTG) residue in position 12 in at least a part of the K-ras genes.

5

Lack of signal from both tubes shows that codon 12 in both alleles is the normal GGT encoding a glycine residue.

10 B. Characterization of the Mutation in Codon 12

The exact characterization of the mutation is done by adding to both tubes a mixture of [^{32}P]-dATP, [^{35}S]-dGTP, [^3H]-dTTP and ddCTP to a final concentration of 1 μM in 15 μl . The [^{32}P]-dATP and the [^{35}S]-dGTP are diluted in unlabelled dATP and dGTP, respectively, to yield similar specific activities as that of the [^3H]-dTTP (about 100 Ci/mmol). The difference in scintillation counting efficiency between the three radioisotopes is taken into account to design a reaction mixture, which will give equal cpm values in the channels used for measurement (see below).

One unit of T7 DNA polymerase is added and the reaction is allowed to proceed for 5 min. at 37°C. The particles are washed and the reaction products are eluted as described in Example 1.

The radioactivity emitted by [^3H], [^{35}S] and [^{32}P] in the eluted product is measured simultaneously in a scintillation counter. In a Rackbeta 1219 counter (Pharmacia/Wallac) the following window settings are used: for measurement of [^3H]: channels 10-90, for measurement of [^{35}S] channels 95-145 and for measurement of [^{32}P] channels 170-220. Before interpretation of the result corrections for the overflow of signal from the [^{35}S] to the [^3H] channels

(24 %) and from the [^{32}P] to the [^{35}S] channels (13 %) are done.

5 The results are interpreted as specified in the table below:

Detection step primer	Signal from			Result	
	^3H	^{35}S	^{32}P	Codon 12	Amino acid
R3	+	-	-	GAT	aspartic acid
R3	-	+	-	GCT	alanine
R3	-	-	+	GTT	valine
R3	-	-	-	GGT	glycine
R4	+	-	-	AGT	serine
R4	-	+	-	CGT	arginine
R4	-	-	+	TGT	cystein
R4	-	-	-	GGT	glycine

25 The mutations in codon 12 of the K-ras gene can optionally be determined as described above, but performing the PCR with biotinylated primer R2 and primer R1. As detection step primer is then used a primer complementary to the opposite strand of the K-ras gene:

R5: 5'- AAC TTG TGG TAG TTG GAG CT (14 - 33) [SEQ.ID.NO. 20]
 R6: 5'- ACT TGT GGT AGT TGG AGC TG (15 - 34) [SEQ.ID.NO. 21]

Example 9

Detection of point mutations in the sequence
encoding codon 12 in the N-ras gene in the presence of
5 non-mutated cells

Synthesis of oligonucleotides

10 Two PCR primers, denoted X1 and X2, respectively,
were synthesized and the primer X1 is biotinylated as
described in Example 1. For detection of the mutation of the
second nucleotide in codon 12 (G replaced by A) a detection
step primer X3 is synthesized. The sequence and position (as
15 nucleotide numbers) of the oligonucleotides on the first exon
of the N-ras gene is given below:

X1: 5'-GAC TGA GTA CAA ACT GGT GG (3 - 22) [SEQ.ID.NO. 22]
X2: 5'-CTC TAT GGT GGG ATC ATA TT (111-91) [SEQ.ID.NO. 23]
20 X3: 5'-ACT GGT GGT GGT TGG AGC AG (15 - 34) [SEQ.ID.NO. 24]

The DNA samples

A cell line known to harbour a nucleotide transition
25 (G → A) in the second position of codon 12 of the N-ras gene
(cell line PA-1, ATCC CRL1572) was used as model system to
mimic a situation where cell samples from AML patients, could
be analyzed for minimal residual mutated cells during follow
up of the treatment of the patients.

30

DNA from PA-1 cells and from normal human
lymphocytes was extracted by standard methods as in Example 7
and mixed to yield a series of samples representing 100% to
0.1% of mutated cells.

Polymerase chain reaction-Amplification

The extracted DNA (100 ng per sample) was amplified using primers X1 and X2 at conditions described in Example 3.

5

Affinity-capture on streptavidin-coated magnetic particles and Annealing of the Detection Step Primers

Eighty μ l of the amplification mixture and 20 μ l of 1 M NaCl was added to 300 μ g of streptavidin-coated magnetic polystyrene beads (Dynabeads^R M-280, Streptavidin, Dynal AS). The samples were kept for 30 min at 20°C, the beads were separated from the reaction mixture using a magnetic particle concentrator (MPC-E, Dynal AS) and the mixture was discarded.

15 The beads were washed three times with 1 ml of 0.5 M NaCl in 20 mM sodium phosphate buffer, pH 7.5, 0.1% Tween 20 and treated twice with 100 μ l of 50mM NaCl, 20 mM MgCl₂, 40 mM Tris-HCl, pH 7.5, containing 2 pmol of detection step primer, X3. The primer was allowed to anneal to the DNA template for

20 10 min at 37°C. One μ l of 0.1 M dithiotreitol, 15 pmol of 3H-labelled dTTP and 1 unit of T7 DNA-polymerase (United Stated Biochemical Corporation) were added to yield a final volume of 15 μ l. The reaction was allowed to proceed for 5 min at 37°C. The beads were washed 3 times and treated with

25 100 μ l of 50 mM NaOH, 0.15 M NaCl for 5 min at 20°C. The eluted radioactivity was measured in a liquid scintillation counter. The results are given in the table below:

Detection of minority point mutations in PA-1 cells

Mutated cells (%)	³ H incorporated (cpm)*
50	23,000
5	3,400
0.5	800
0	490

5

10

*) Incorporation of ³H corresponding to A in codon 12 (GAT)

Conclusion

15

The result shows that the method according present invention detects as littes as 0.25% of mutated monoallelic N-ras DNA from a background of normal DNA. This sensitivity is well in the range needed to screen AML patients and to monitor the identified mutations during treatment and follow up of the disease.

25

Example 10

Detection of point mutation (A - T) in the sequence encoding codon 215 (amino acid sequence number is relative to the NH2 terminal proline) of HIV-1 reverse transcriptase gene

30

Synthesis of oligonucleotides

35

Two PCR primers (H1 and H2) and one detection step primer (H3) are synthesized by the method described in

Example 1. The primer H2 is biotinylated as described in Example 1. The nucleotide sequences of the oligonucleotides and their location on the HIV-1 reverse transcriptase gene (nucleotide numbering is according to Ratner et al., 1985, Nature 313: 277-281) are the following:

H1: 5' -GAG AAT CCA TAC AAT ACT CCA (2286-2306) [SEQ.ID.NO. 25]
H2: 5' -TAA CCC ATC CAA AGG AAT GGA (2824-2804) [SEQ.ID.NO. 26]
H3: 5' -ATC TGT TGA GGT GGG GAC TT (2752-2771) [SEQ.ID.NO. 27]

10

Polymerase chain reaction-Amplification and Affinity-capture

The DNA is extracted from HIV-1 infected cells as described in Example 1. The DNA (100 ng per sample) is amplified with the primers H1 and H2 (final concentration 0.2 mM) as described in Example 1 with the exception that the primer annealing temperature is 50° C. The biotinylated amplified DNA fragments are captured on streptavidin-coated microtiter plate wells as described in Example 4. Each sample is bound to two parallel wells. The DNA fragments bound to the microtiter plate walls are denatured and washed as described in Example 4.

25

Identification of the A - T mutation in codon 215

The annealing of the detection primer H3 and the detection reaction are done simultaneously at 50° C for 10 min in 50 mM Tris-HCl, pH 8.8, 1.5 mM MgCl₂, 15 mM (NH₄)₂SO₄, 0.1 % Tween 20, 0.01 % gelatin and 0.2 μM primer H3 containing for A reaction: 1 μCi of ³H-dATP (82 Ci/mmol, Amersham, UK), 0.8 μM of ddTTP and ddCTP and 1 U of Taq DNA polymerase (Promega). For T-reaction: 1 μCi of ³H-dTTP (98 Ci/mmol) and 0.8 μM of ddATP and ddCTP. The wells are washed three times with 0.1 % Tween 20 in TBS and the

35

radioactivity is eluted with 60 ml of 50 mM NaOH for 5 min at room temperature. The eluted radioactivity is measured in liquid scintillation counter.

5

Interpretation of results

A positive signal from only the A-reaction shows that the virus isolate is wild type. A positive signal from only the T-reaction shows that the virus isolate carries the mutation of the first nucleotide of codon 215. Consequently, the amino acid Thr has changed to Phe or Tyr. Positive reactions from both reactions show that the virus is a mixed population of wild type and mutant viruses.

15

The variable nucleotides in codons 67 and 70 could be determined in a similar way from the same amplification product.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Orion-yhtymä Oy
- (ii) TITLE OF INVENTION: Method and Reagent for Determining Specific Nucleotide Variations
- (iii) NUMBER OF SEQUENCES: 27
- (iv) CORRESPONDENCE ADDRESS:
- Orion Corporation
ORION PHARMACEUTICA,
Patent Department
P.O.Box 65
02102 ESPOO
FINLAND

(v) COMPUTER READABLE FORM: <>

- (A) MEDIUM TYPE: <>
(B) COMPUTER: <>
(C) OPERATING SYSTEM: <>
(D) SOFTWARE: <>

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: To be determined
(B) FILING DATE:
(C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 22 bases
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAG GAG TTG AAG GCC TAC AAA T 22

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 22 bases
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAA CAA CTG AGC CCG GTG GCG G 22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCG CGG ACA TGG AGG ACG TG 20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATG CCG ATG ACC TGC AGA AG 20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 22
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCG CGG GCC CCG GCC TGG TAC A 22

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 22
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6

GGA TGG CGC TGA GGC CGC GCT C 22

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTA CTG CAC CAG GCG GCC GC 20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 22
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGC CTG GTA CAC TGC CAG GC 22

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAT TTG CTT CTG ACA CAA CT 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CAT GGT GCA CCT GAC TCC TG 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CAA CTT CAT CCA CGT TCA CC 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAG TAA CGG CAG GCG GCC GC 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 23
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTG GAG CCT TCA GAG GGT AAA AT 23

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 24
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGG CAC CAT TAA AGA AAA TAT CAT 24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 23
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CAT GCT TTA ATG ACG CTT CTA TA 23

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATG ACT GAA TAT AAA CTT GTG 21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTC GTC CAC AAA ATG ATT CTG 21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AAG GCA CTC TTG CCT ACG CCA 21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AGG CAC TCT TGC CTA CGC CAC 21

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAC TTG TGG TAG TTG GAG CT 18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACT TGT GGT AGT TGG AGC TG 18

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GAC TGA GTA CAA ACT GGT GG 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTC TAT GGT GGG ATC ATA TT 20

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACT GGT GGT GGT TGG AGC AG 20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAG AAT CCA TAC AAT ACT CCA 21

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 21
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TAA CCC ATC CAA AGG AAT GGA 21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATC TGT TGA GGT GGG GAC TT 20

CLAIMS

1. A method for detecting a specific nucleotide variation at a defined site in a target nucleic acid polymer wherein a first nucleotide residue is replaced by a second nucleotide residue, comprising
 - (a) hybridising a detectable amount of a target nucleic acid polymer in single-stranded form with an oligonucleotide primer, the detection step primer, comprising a plurality of nucleotide residues, said primer being complementary to the nucleotide sequence of interest in a region disposed toward the 3' end from the defined site such that when the primer is hybridized to the polymer there are no nucleotide residues between the defined site and the 3' end of the primer that are identical to the first or second nucleotide residues to be detected;
 - (b) extending the primer using a polymerizing agent in a mixture comprising one or more nucleoside triphosphates wherein the mixture includes at least one nucleoside triphosphate complementary to either the first or second nucleotide residue which comprises means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer, and optionally one or more chain terminating nucleoside triphosphates; and
 - (c) detecting the incorporation of the nucleoside triphosphate, whereby the identity of the nucleotide residue at the defined site is determined.
2. A method for detecting a plurality of specific nucleotide variations at defined sites in a target nucleic acid polymer wherein at least first nucleotide residue is replaced by a fourth nucleotide residue at a second defined site, comprising
 - (a) hybridizing a detectable amount of a target nucleic acid polymer in single-stranded form with a first detection

step primer, comprising a plurality of nucleotide residues, said primer being complementary to the nucleotide sequence of interest in a region disposed toward the 3' end from the first defined site such that when the primer is hybridized to the polymer there are no nucleotide residue between the first defined site and the 3' end of the primer that are identical to the first and second nucleotide residues;

(b) extending the first detection step primer using a polymerizing agent in a mixture comprising one or more nucleoside triphosphates wherein the mixture includes at least one nucleoside triphosphate complementary to the first or second nucleotide residue which comprises means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer, and optionally one or more chain terminating nucleoside triphosphates;

(c) detecting the incorporation of the nucleoside triphosphate, whereby the identity of the nucleotide residue at the first defined site is determined;

(d) removing the extended first detection step primer formed in step (c) from the target nucleic acid polymer; and

(e) adding a second detection step primer, said primer being complementary to the nucleotide sequence of interest in a region disposed toward the 3' end from the second defined site such that when the primer is hybridized to the immobilized polymer there are no nucleotide residues between the second defined site and the 3' end of the primer that are identical to the third or forth nucleotide residues to be detected.

3. A method of detecting in a patient a predisposition to a genetic disorder resulting from a specific nucleotide variation at a defined site in genetic material of the patient, wherein a first nucleotide residue is replaced by a second nucleotide residue, comprising the steps of

(a) obtaining a sample containing a detectable amount of genetic material derived from the patient;

(b) hybridizing the detectable amount of genetic material in a single-stranded form with a first oligonucleotide primer, the first detection step primer, comprising a plurality of nucleotide residues, said primer being complementary to the nucleotide sequence of interest in a region disposed toward the 3' end from the first defined site such that when the primer is hybridized to the genetic material there are no nucleotide residues between the defined site and the 3' end of the primer that are identical to the first and second nucleotide residues;

(c) extending the primer using a polymerizing agent in a mixture comprising one or more nucleoside triphosphates wherein the mixture includes at least one nucleoside triphosphate complementary to either the first or second nucleotide residue which comprises means for detecting the incorporation of the nucleoside triphosphates in a nucleic acid polymer, and optionally one or more chain terminating nucleoside triphosphate; and

(d) detecting the incorporation of the nucleoside triphosphate, whereby the identity of the nucleotide residue at the defined site, and thus whether the patient has a predisposition for the associated genetic disorder is determined.

4. A method for detecting the existence of point mutations at a defined site in the genome of a microorganism leading to altered pathogenicity or resistance to therapy in the microorganism, wherein a first nucleotide is replaced by a second nucleotide residue, comprising the steps of

(a) obtaining as sample containing a detectable amount of genetic material derived from the microorganism;

(b) hybridizing the detectable amount of genetic material in a single-stranded form with an oligonucleotide primer, the detection step primer, comprising a plurality of nucleotide residues, said primer being complementary to the nucleotide sequence of interest in a region disposed toward

the 3' end from the defined site such that when the primer is hybridized to the genetic material there are no nucleotide residues between the defined site and the 3' end of the primer that are identical to the first or second nucleotide residues to be detected;

(c) extending the primer using a polymerizing agent in a mixture comprising one or more nucleoside triphosphates wherein the mixture includes at least one nucleoside triphosphate complementary to either the first or second nucleotide residue which comprises means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer, and optionally one or more chain terminating nucleoside triphosphates; and

(d) detecting the incorporation of the nucleoside triphosphate, whereby the identity of the nucleotide residue at the defined site and thus whether a point mutation has occurred is determined.

5. A method for detecting cells having a point mutation at a defined site in the genetic material, wherein a first nucleotide residue is replaced by a second nucleotide residue, when said mutated cells are mixed in a cell population, comprising the steps of:

(a) obtaining a detectable quantity of genetic material from the cell population while maintaining the ratio of mutated to unmutated cells;

(b) hybridizing the detectable amount of genetic material in a single-stranded form with an oligonucleotide primer, the detection step primer, comprising a plurality of nucleotide residues, said primer being complementary to the nucleotide sequence of interest in a region disposed toward the 3' end from the defined site such that when the primer is hybridized to the genetic material there are no nucleotide residues between the defined site and the 3' end of the primer that are identical to the first or second nucleotide residues to be detected;

(c) extending the primer using a polymerizing agent in a mixture comprising one or more nucleoside triphosphates wherein the mixture includes at least one nucleoside triphosphate complementary to either the first or second nucleotide residue which comprises means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer, and optionally one or more chain terminating nucleoside triphosphates; and

(d) detecting the incorporation of the nucleoside triphosphate, whereby the identity of the nucleotide residue at the defined site and thus whether mutated cells are present is determined.

6. A method according to claims 1 - 5, further comprising the step of immobilizing the target nucleic acid polymer to a solid support prior to step (a).
7. A method according to claims 1 - 5, wherein the primer is complementary to a region of the nucleotide sequence of interest extending toward the 3' end of the target nucleic acid polymer from the nucleotide residue immediately adjacent to the defined site.
8. A method according to claims 1 - 5, wherein the nucleoside triphosphate comprising means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer is a deoxynucleoside triphosphate.
9. A method according to claims 1 - 5, wherein the nucleoside triphosphate comprising means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer is a dideoxynucleoside triphosphate.
10. A method according to claims 1 - 5, wherein the mixture includes a second nucleoside triphosphate comprising a second means, different from said first means, for detecting the

- incorporation of the second nucleoside triphosphate in a nucleic acid polymer.
11. A method according to claim 2, wherein the extended product of step (d) is eluted before determining the incorporation of the incorporated nucleoside triphosphate.
 12. A method according to claim 2, wherein the nucleotide variations are detected in one single step by adding a plurality of detection step primers and differently labelled nucleoside triphosphates identifying the variable nucleotide residues.
 13. A method according to claims 1 - 5, wherein the detectable amount of target nucleic acid polymer is obtained by performing a modified amplification reaction wherein at least one amplification primer comprises a first attachment moiety bonded to the primer.
 14. A method according to claim 4, wherein the microorganism is HIV.
 15. A method according to claim 14, wherein the point mutation is at a site selected from among Asp 67, Lys 80 and Thr 215.
 16. A method according to claim 5, wherein the cells are lymphocytes.
 17. A method according to claim 16, wherein the cells are leukemic cells.
 18. A kit for use in determining specific nucleotide variations in a target nucleic acid polymer comprising in packaged combination
 - (a) at least one amplification primer comprising an oligonucleotide which is complementary to and hybridizes with

a portion of the target nucleic acid polymer and which is effective as a primer for enzymatic nucleic acid polymerization and a first attachment moiety;

(b) at least one detection step primer comprising an oligonucleotide which is complementary to and hybridizes with a portion 3' to a variable nucleotide of the target nucleic acid polymer; and optionally

(c) at least one solid support comprising a solid matrix and at least one attachment site which is capable of immobilizing the oligonucleotide of the amplification probe through the first attachment moiety; and

(d) at least one nucleoside triphosphate containing means for detecting the incorporation of the nucleoside triphosphate in a nucleic acid polymer.

19. A kit according to claim 18 for use in the identification of the nucleotide variation of apolipoprotein E polymorphism, wherein the detection step primer comprises the sequence 5'-GCG CGG ACA TGG AGG ACG TG.
20. A kit according to claim 18 for use in the identification of the nucleotide variation of apolipoprotein E polymorphism, wherein the detection step primer comprises the sequence 5'-ATG CCG ATG ACC TGC AGA AG.
21. A kit according to claim 18 for use in the identification of the nucleotide variation of apolipoprotein E polymorphism, wherein the detection step primer comprises the sequence 5'-GTA CTG CAC CAG GCG GCC GC.
22. A kit according to claim 18 for use in the identification of the nucleotide variation of apolipoprotein E polymorphism, wherein the detection step primer comprises the sequence 5'-GGC CTG GTA CAC TGC CAG GC.
23. A kit according to claim 18 for use in the detection of the

- nucleotide variation in codon 6 of the human β -globin gene causing sickle cell anemia, wherein the detection step primer comprises the sequence 5'- CAT GGT GCA CCT GAC TCC TG.
24. A kit according to claim 18 for use in the detection of the nucleotide variation in codon 6 of the human β -globin gene causing sickle cell anemia, wherein the detection step primer comprises the sequence 5'- CAG TAA CGG CAG GCG GCC GC.
25. A kit according to claim 18 for use in the detection of a nucleotide variation in codon 12 of the K-ras gene, wherein the detection step primer comprises the sequence 5'- AAG GCA CTC TTG CCT ACG CCA.
26. A kit according to claim 18 for use in the detection of a nucleotide variation in codon 12 of the K-ras gene, wherein the detection step primer comprises the sequence 5'- AGG CAC TCT TGC CTA CGC CAC.
27. A kit according to claim 18 for use in the detection of a nucleotide variation in codon 12 of the K-ras gene, wherein the detection step primer comprises the sequence 5'- AAC TTG TGG TAG TTG GAG CT.
28. A kit according to claim 18 for use in the detection of a nucleotide variation in codon 12 of the K-ras gene, wherein the detection step primer comprises the sequence 5'- ACT TGT GGT AGT TGG AGC TG.
29. A kit according to claim 18 for use in the detection of a nucleotide variation in codon 12 of the N-ras gene, wherein the detection step primer comprises the sequence 5'- ACT GGT GGT GGT TGG AGC AG.
30. A kit according to claim 18 for use in the detection of resestancy to AZT in HIV-1 viruses, wherein the detection

step primer comprises the sequence 5'- ATC TGT TGA GGT GGG GAC TT.

31. A kit according to claim 18 for use in the detection of cystic fibrosis, wherein the detection step primer comprises the sequence 5'- TGG CAC CAT TAA AGA AAA TAT CAT.
32. A reagent for detecting the presence of a point mutation in which a normal nucleic acid residue is replaced by an abnormal nucleic acid residue at a defined site within a gene of interest, comprising an oligonucleotide of sufficient length to act as a primer for an enzyme catalyzed chain extension nucleic acid polymerization reaction, said oligonucleotide primer having a sequence which is complementary to a region of the gene of interest beginning with the nucleotide residue immediately adjacent to and toward the 3' end of the gene from the defined site and extending away from the defined site toward the 3' end of the gene whereby enzyme catalyzed chain extension nucleic acid polymerization will commence by adding a nucleic acid residue complementary to either the normal nucleic residue or the abnormal nucleic acid residue.
33. A reagent according to claim 32, wherein the oligonucleotide has a length of from 10 - 40 nucleotide residues
34. A reagent according to claim 32 having the sequence 5'- GCG CGG ACA TGG AGG ACG TG.
35. A reagent according to claim 32 having the sequence 5'- ATG CCG ATG ACC TGC AGA AG.
36. A reagent according to claim 32 having the sequence 5'- GTA CTG CAC CAG GCG GCC GC.
37. A reagent according to claim 32 having the sequence 5'-

GGC CTG GTA CAC TGC CAG GC.

38. A reagent according to claim 32 having the sequence 5'-
CAT GGT GCA CCT GAC TCC TG.
39. A reagent according to claim 32 having the sequence 5'-
CAG TAA CGG CAG GCG GCC GC.
40. A reagent according to claim 32 having the sequence 5'-
AAG GCA CTC TTG CCT ACG CCA.
41. A reagent according to claim 32 having the sequence 5'-
AGG CAC TCT TGC CTA CGC CAC.
42. A reagent according to claim 32 having the sequence 5'-
AAC TTG TGG TAG TTG GAG CT.
43. A reagent according to claim 32 having the sequence 5'-
ACT GGT GGT GGT TGG AGC AG.
44. A reagent according to claim 32 having the sequence 5'-
ATC TGT TGA GGT GGG GAC TT.
45. A reagent according to claim 32 having the sequence 5'-
TGG CAC CAT TAA AGA AAA TAT CAT.

Detection

Step Primer: 5'- Y Y Y Y Y Y Y Y Y

Immobilized

Target: 3'- X X X X X X X X X₁X'X X X X - attachment
X₂ moiety

Added nucleoside triphosphate: a) ddY₁* or ddY₂*
b) ddY₁* and ddY₂⁺
c) dY₁ (and ddY')
d) dY₁* and dY₂⁺

FIG. 1

Detection

Step Primer: 5'- Y Y Y Y Y Y Y Y

Immobilized

Target: 3'- X X X X X X X_n X₁ X' X" X X X - attachment
moiety

Added nucleoside triphosphate: a) dY_n and (ddY₁* or ddY₂*),
only if X_n is not X₁ or X₂

FIG. 2

Detection

Step Primers: 5'- Y Y Y Y Y Y Y
5'- Y Y Y Y Y Y Y Y₁
5'- Y Y Y Y Y Y Y Y Y₂

Immobilized

Target: 3'- X X X X X X X X₁ X₃ X X X X - attachent
X₂ X₄ moiety

Added nucleoside triphosphate: Step 1: ddY₁* and/or ddY₂*
Step 2: ddY₃* and/or ddY₄*

FIG. 3